

Exploring a Framework for Consequential Validity for Performance-Based Assessments

A DISSERTATION  
SUBMITTED TO THE FACULTY OF THE  
UNIVERSITY OF MINNESOTA  
BY

Su Jung Kim

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Dr. Mistilina Sato, Adviser

August 2017



## **Acknowledgement**

I would first like to thank my committee members, Dr. Barbara Billington, Dr. Mistilina Sato, Dr. Stacy Ernst, and Dr. Stuart Yeh. Every step to complete my PhD came with its challenges and frustrations. Collectively you all made this process as valuable and enjoyable as it could be. You guided, mentored, and importantly challenged me when it needed. I really appreciate your supports and care. I will take my appreciation for all of you wherever I move forward.

The door to Dr. Sato office was always open whenever I had questions and issues with my dissertation study and/or writing. She consistently allowed this paper to be my own work, but guided me in the right the direction whenever she thought I needed it. It was an honor to be your advisee for the last seven years. You are my role model as a teacher educator.

I would also like to thank the experts who were involved in the validation survey for this research project: Dr. Andrea Whittaker, Dr. Irena Nayfeld, and Dr. Raymond Pecheone, at Stanford University. With your passionate input, the validation survey could not have been successfully conducted.

I would like to thank you - my friends and colleagues whom I have met in the CEHD community. The MnEDS group – Miranda, Jehanne, Jessica, Jenna, Justin, and Amanda, you all provided me shelter, friendship, and laughter. The TERC group – you are all risk-takers and innovators. It was a wonderful experience to learn from and teach each other.

I want to acknowledge and thank the students, administrators and faculty of the department of Curriculum and Instruction.

## **Dedication**

This work is dedicated to my parents. Mom and Dad – both of you provided your unfailing support and continuous encouragement throughout my years of study in the United States the last 14 years and through the process of completing my PhD journey. It would have been impossible to achieve this accomplishment without you. I am the luckiest girl in the world and I am so honored to be your daughter. I love you both with all my heart.

## **Abstract**

This study explores a new comprehensive framework for understanding elements of validity, specifically for performance assessments that are administered within specific and dynamic contexts. The adoption of edTPA is a good empirical case for examining the concept of consequential validity because this assessment has been implemented at the state level over several years in Minnesota, adding a much needed large-scale case to the literature on situated assessment validity (Moss, 1996). By drawing on the work of several measurement theorists (Frederiksen & Collins, 1989; Linn, Baker, & Dunbar, 1991; Uhlenbeck, Verloop, & Beijaard, 2002), I have constructed a framework of six dimensions for exploring the concept of consequential validity: 1) educational consequence, 2) meaningfulness, 3) directness, 4) transparency, 5) fairness, and 6) usability. This framework serves both as an analytic frame for data analysis and as our effort to further define and synthesize what is meant by consequential validity within the measurement and assessment field. This study used three data collection strategies: individual and group interviews with teacher candidates and teacher educators and a survey of teacher educators. 23 teacher candidates and 11 teacher educators were interviewed and they represented the edTPA participants at the University of Minnesota. In addition, 22 teacher educators from other institutions participated in a survey across Minnesota. Findings from this study indicate that teacher candidates and teacher educators identify both positive and negative consequences of edTPA based on its design and expectations. However, the consequences of edTPA were strongly affected by local implementation contexts. By using the six dimensions within the proposed framework,

this study uncovers links between consequences and contexts. This suggests that implementation of a new assessment that is meant to be educative may also rely on developing a particular “assessment culture” in which data interpretation and use is a more commonly accepted practice.

## Table of Contents

Acknowledgement .....	i
Dedication .....	iii
Abstract .....	iv
List of Tables .....	viii
List of Figures .....	ix
CHAPTER 1. INTRODUCTION .....	1
Problem Statement .....	2
Research Questions .....	5
Overview of Chapters .....	6
CHAPTER 2. REVIEWS OF THE LITERATURE .....	8
Performance-Based Assessments in Teaching and Teacher Education.....	10
Characteristics of authentic assessments for teachers.....	11
Concerns about performance assessments for teaching.....	12
Establishing Validity of Performance-based Assessments .....	13
Content representation .....	14
Construct alignment with external standards.....	15
Scoring reliability.....	16
Fairness and bias studies .....	16
Underlying Validity Principles in Empirical Validity Studies .....	17
Traditional view of validity .....	17
Messick's unified view of construct validity .....	19
Alternative approaches to examining validity of performance-based assessment	20
Moss's situated inquiry for establishing validity .....	21
Multiple dimensions of performance-based assessments .....	23
Educational Consequences.....	24
Meaningfulness.....	24
Directness.....	25
Transparency.....	25
Fairness.....	26
Usability.....	26
CHAPTER 3. METHODS .....	29
Research Questions .....	30
Research Design.....	31
Data Collection Methods .....	33
Institution selection .....	34
Participant selection .....	35
Data Analysis Methods .....	40
CHAPTER 4. RESEARCH FINDINGS.....	44
Teacher Candidates' Perspectives.....	45



Educational Consequences.....	45
Meaningfulness.....	49
Transparency.....	52
Directness.....	55
Fairness.....	56
Usability.....	57
Teacher Educators' Perspectives .....	59
Educational Consequences.....	60
Meaningfulness.....	65
Transparency.....	70
Directness.....	74
Fairness.....	80
Usability.....	85
Summary of Findings.....	90
CHAPTER 5. DISCUSSION AND FUTURE RESEARCH.....	94
Six Dimensions of Consequential validity.....	95
Educational Consequences.....	95
Meaningfulness.....	97
Transparency.....	99
Directness.....	101
Fairness.....	102
Usability.....	103
Implications.....	105
Implications for teacher education programs.....	105
Implications for validity research studies .....	106
Limitations to this study .....	107
Conclusions.....	109
Bibliography .....	112
Appendix A.....	117
Appendix B.....	118
Appendix C .....	120

## List of Tables

Table 3.1 <i>Timeline of data collection activities</i> .....	33
Table 3.2 <i>Teacher candidate participants at the UMN and their licensure program</i> .....	36
Table 3.3 <i>Teacher educator participants at the UMN and their licensure program</i> .....	38
Table 3.4 <i>Survey completers and their institutional background</i> .....	40
Table 4.1 <i>Survey results in educational consequences at classroom level</i> .....	61
Table 4.2 <i>Survey results in educational consequences at program/institutional level</i> .....	64
Table 4.3 <i>Survey results in meaningfulness at classroom level</i> .....	66
Table 4.4 <i>Survey results in meaningfulness at program/institutional level</i> .....	68
Table 4.5 <i>Survey results in transparency at classroom level</i> .....	71
Table 4.6 <i>Survey results in transparency at program/institutional level</i> .....	73
Table 4.7 <i>Survey results in directness at classroom level</i> .....	76
Table 4.8 <i>Survey results in directness at program/institutional level</i> .....	78
Table 4.9 <i>Survey results in fairness at classroom level</i> .....	80
Table 4.10 <i>Survey results in fairness at program level</i> .....	83
Table 4.11 <i>Survey results in fairness at institutional level</i> .....	84
Table 4.12 <i>Survey results in usability at classroom level</i> .....	86
Table 4.13 <i>Survey results in usability at program/institutional level</i> .....	88

## **List of Figures**

<i>Figure 1.</i> Modified Data Analysis Model From Miles and Huberman's (1994).....	41
---	----

## CHAPTER 1

### INTRODUCTION

Performance-based assessments are broadly used for formative and summative purposes in all education settings. When this type of assessment provides multiple learning opportunities to apply knowledge and skills being measured and to revise their work, learners deepen their understanding in knowledge and skills and teachers gain information for instructional decisions. Also, performance-based assessment can be used for summative decisions about learners' understanding of knowledge and skills in particular domains.

With a special interest in measuring and supporting teacher candidates' ability to perform teaching, teacher preparation programs have been increasingly adopting performance-based assessments in extensive and intensive clinical experience. Along with this trend, a growing number of research studies have found positive learning experiences for teacher candidates who engage in these assessments. Various formats and designs of performance-based assessments have been developed and implemented. Among them, the Educative Teacher Performance-based Assessment (edTPA) is the first initiative to use a uniform assessment for teacher candidates across states.

Currently, edTPA has been implemented in more than 600 teacher education programs since 2013 Fall and 41 states and the District of Columbia are participating in edTPA (Stanford Center for Assessment, Learning, and Equity (SCALE)). This assessment asks teacher candidates to create a digital portfolio that includes written reflections and videos of their teaching with K-12 students. It is the first initiative to

administer a national level assessment in teacher education across the United States. State departments of education have made state-specific decisions for adopting this new assessment into their existing teacher education system.

Beyond the successful implementation of the first national level performance-based assessment for teacher candidates, practitioners from the field have raised several questions. One big question that I am interested in pursuing is how a uniform assessment is determined to be a valid measure of teacher candidate performance when those candidates are educated in different ways across multiple institutions and complete this assessment in a diverse array of classrooms. This concern directly relates to a broader discussion about how we determine the validity of performance assessments.

This chapter provides an overview of this study. It starts with a problem statement to address an issue relate to uses of performance-based assessment. By proposing the problem statement, it guides to establish an arena of the research study. In addition, research questions are introduced to inform a focus of this study. Lastly, a summary of each chapter is followed at the end of this chapter.

### **Problem Statement**

A series of studies has evaluated the validity of edTPA during the early phase of implementation (SCALE, edTPA Administrative Report, 2013). These studies established evidence that shows how edTPA is a valid measure to capture teaching performance by examining performance tasks and comparing the assessment with professional teaching standards. The studies conclude that tasks in edTPA are well developed and resemble real teaching practices. Also, they map out alignments between

edTPA and other professional standards including Interstate New Teacher Assessment and Support Consortium (InTASC) Model Core Standards and National Board for Professional Teaching Standards (NBPTS).

The series of studies provide strong evidence of how edTPA is well designed to measure what is intended. However, these studies did not intend to evaluate local implementation of edTPA. Practitioners have expressed doubts about how the results of edTPA accurately reflect the performance of teacher candidates in uncontrollable and unpredictable assessment sites which seemingly have obvious influence on the assessment activities and results.

The current edition of the Standards of Testing and Assessments (American Psychological Association, American Educational Research Association, & National Council on Measurement and Education (AERA, APA, & NCME), 2014) offers general guidelines to uphold the quality of all assessments. Validity and reliability are core components to evaluate all tests and assessments. Also, the standards hold a strong viewpoint that all assessments, including performance-based assessments, must be evaluated by the approach outlined in the standards. However, this dominant viewpoint sits in tension with implementation of more complex performance-based assessments.

Validity theorists agree that it is fundamental to fully explore characteristics of an assessment to test the validity of an assessment properly. A distinctive characteristic of performance-based assessments in teaching and teacher education generally is to provide professional development opportunities by situating the assessment in real-world teaching. In the case of edTPA, it is designed to scaffold teacher candidates' learning to

teach during their clinical experience with context-specific feedback. Also, teacher educators learn how to improve pedagogical practices in teacher education programs by using edTPA candidate performances and results. Overall, it is designed to create meaningful and significant impact on teacher candidates, teacher educators, and teacher education programs. Yet, the current edTPA validity studies do not fully explore how learning experiences directly or indirectly interact with local contexts.

As concepts of assessment validity have evolved in the last few decades, it is important that new definitions and guidelines for validity support an accurate and complex view of the implementation of performance-based assessments in local contexts. We may need to reconsider the current stance that the assessment and measurement field holds toward examining validity of performance-based assessments. With the recognition of this tension, this study examined the formative and educative construct of performance-based assessments within a validity framework by using a case of the edTPA.

Based on reviews of literature and professional conversations with teacher educators, I have identified a set of critical issues related to the implementation and use of edTPA. First, there is a gap between findings from completed validity studies and concerns raised from practitioners. It is urgent to address this gap and provide guidelines that better support implementation practices. Second, the current scope of validity does not consider the dynamic contexts as evidence to validate assessments with characteristics of performance-based assessments. This viewpoint should be reconsidered with respect to practical advice for practitioners. Finally, a concept of *consequential*

*validity* could be relevant to understand educational outcomes of performance-based assessment. This concept is still under development in the field and I try to further our understanding of consequences during assessment adoption and implementation in this study.

### **Research Questions**

In this study, I explore the concept of consequential validity as one aspect of construct validity of an assessment. The key evidence of consequential validity is located in specific contexts and processes of implementation of assessments which have impacts on educational outcomes. In order to explore this concept, I have theorized a framework for consequential validity based on an expansive literature review. This study tests the usefulness and viability of this framework within the context of the edTPA implementation in the state of Minnesota from 2013-2014. My research questions reflect this conceptual exploration within an empirical context by posing both questions about consequential validity and edTPA implementation as follows:

1. What constructs should be included in a framework that describes the consequential validity of an assessment?
2. How can a consequential validity framework help describe the implementation of a standardized complex performance assessment administered in local contexts?

In order to address these questions about consequential validity, I will pursue questions related to the implementation of the edTPA in Minnesota. The implementation questions I ask include 1) how was the edTPA implemented locally?; 2) how did the implementation of edTPA influence new decisions about program policy, curriculum, and



classroom for teacher candidates?; and 3) how are edTPA results used by instructors, programs, and institutions. By pursuing these questions, I gathered evidence to test and refine a framework for consequential validity.

The goal of this study is to further our understanding of how the validity of a performance assessment interacts with local contexts as a process of assessment validity. By naming the consequence and describing the local implementation practices and barriers, I think this study will be helpful in implementation of performance-based assessment in real-world settings.

### **Overview of Chapters**

Chapter two reviews the current research studies to understand the main perspectives on validity of performance-based assessment. The reviewed literature reveals critical considerations for evaluating the validity of performance-based assessments. With a special focus on educational consequences of performance-based assessments, a concept of consequential validity is defined and developed throughout this study.

Chapter three introduces the research design, data collection, and data analysis methods. This study is based on a mixed-methods design drawing on both survey and interview data with both teacher candidates and teacher educators.

Chapter four provides a summary of findings in this study and addresses research question two. Findings from two groups of participants are analyzed across six dimensions of a consequential validity framework including educational consequences, meaningfulness, directness, transparency, fairness, and usability.

Chapter five addresses research question one by summarizing how a concept of consequential validity can support our understanding of the implementation of performance-based assessments and examining the strengths and limitations of the consequential validity framework I use in the analysis. The summary connects to implications of assessment implementation and implications of establishing validity of performance-based assessments.

## CHAPTER 2

### REVIEWS OF THE LITERATURE

A growing number of studies have recognized educative purposes of performance-based assessments for teachers' professional development (Darling-Hammond & Snyder, 2000; Kennedy, 2010; Sato, Wei, Darling-Hammond, 2008). For example, Sato and her colleagues found that teacher's classroom instruction can be changed and improved through professional learning opportunities embedded in the National Board for Professional Teaching Standards' portfolio assessment. To this end, numerous performance-based assessments have been developed and used in different educational settings to provide learning opportunities.

On the other hand, some practitioners and users of performance-based assessments raise questions about the quality and usefulness of performance-based assessments beyond educative purposes. With recognition of these concerns, a series of validity studies of performance-based assessments have been conducted (Wei & Pechione, 2010). Most of these validity studies were conducted based on the Standards for Educational and Psychological Testing (APA, AERA, & NCME, 1999, 2014) which is grounded in Messick's unified view of validity (Messick, 1995). The central and official definition of validity (APA, AERA, & NCME, 1999, 2014) is drawn from Messick's key definition of validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment" (Messick, 1990, p. 5). The standards aim to offer practical guidelines for test

developers and users to evaluate and improve the quality of educational assessments (Shepard, 1993).

To specify the practical guidelines and to have practical validation processes, theorists have proposed various ideas and schemes to establish validity with different emphases and special interests. Messick (1989) took a lead to expand a definition boundary of validity by suggesting consequences as validity evidence. Following this, Green (1998) attempted to integrate evidence of test-users' behaviors during assessment activities as a source of validity evidence (Green, 1998). These attempts were criticized by Popham (1991) because loosely defined validity and its practice may confuse practitioners. These attempts opened the field of test validity to consider establishing practical and useful approaches to evaluate the quality and usefulness of assessments.

In this chapter, I review characteristics of performance-based assessment as authentic assessments in the field of teaching and teacher education. This review contributes to our understanding of key constructs of performance-based assessments to establish their validity. My examination of completed validity studies of performance-based assessments in the field reveals a gap in how to validate performance-based assessments. It reveals a need to explore alternative perspectives on validity which are relevant to assessments situated in dynamic and fluid contexts of performance-based assessment. Lastly, a set of alternative perspectives for performance-based assessments is reviewed and suggested as a framework for this study.

## **Performance-Based Assessments in Teaching and Teacher Education**

Performance-based assessments use practical tasks that help determine what students know and how they apply their knowledge. During such assessments, students demonstrate knowledge and its application rather than select an answer from a pre-determined list. Often, some large-scale assessments such as the Scholastic Aptitude Test (SAT), American College Testing (ACT), and Advanced Placement exams (AP) include both a set of multiple-choice questions and a portion of performance-based tasks. Advocates of performance-based assessment believe that “because they require students to actively demonstrate what they know, performance-based assessments may be a more valid indicator of students’ knowledge and abilities” (Sweet, 1993).

Researchers in the field of assessment have argued for the benefits of using performance-based assessments for both students and teachers. Performance-based assessments ask the test-takers to demonstrate what they have learned through a variety of tasks which can provide the assessor with more information than multiple-choice tests. Darling-Hammond and Adamson (2010) argued, “well-designed performance assessments yield a more complete picture of students’ abilities and weakness, and can overcome some of the validity challenges of assessing English language learners and students with disability” (p.3). Similarly, Edward Haertel noted, “these new forms of assessment would promote active engagement both in learning and in demonstrating what had been learned. They would serve as models of sound instructional activities...As the line between teaching and testing blurred, classroom time would be better employed” (Haertel, 1999, p. 663). These studies found that when assessment processes and

instruction are integrated in classrooms, it is a powerful approach to enhance learning. Assessment processes can provide richer information about learners by showing their ability to apply knowledge in practice. Since scores from on-demand assessments are limited in how they provide information about why learners choose correct or wrong answers, performance artifacts as a part of the assessment process are valuable information for guiding instruction for learners. A growing number of research studies have recognized these positive effects of performance-based assessments in teaching and teacher education.

### **Characteristics of authentic assessments for teachers**

Darling-Hammond and Snyder (2000) identified four characteristics of authentic assessments for enhancing learning to teach. First, authentic assessments involve sampling the actual knowledge, skills, and dispositions desired of teachers in real teaching and learning contexts. Second, the actual knowledge, skills, and dispositions should be integrated into teaching practices to demonstrate multiple facets of teaching. Third, multiple sources of evidence need to be collected over time and in diverse contexts. Finally, to provide more helpful advice for teachers, the collected evidence in performance-based assessments should be evaluated by relevant experts against criteria or standards that matter for the performance of teaching. These four characteristics directly point to performance-based assessment as authentic assessment of teaching performance.

Particularly within teacher education contexts, Darling-Hammond and Snyder (2000) suggested considering an additional requirement for authentic assessment that

provides multiple opportunities for learning and practicing the desired outcomes and for feedback and reflection. This new expectation is closely related to the core function of teacher education programs, which aim to increase teacher candidates' ability to reflect on and learn from their teaching practices. Authentic assessments for teacher candidates should include opportunities for learning from feedback and reflection that both support the development of greater levels of competence and measure critical attributes of an effective teacher such as the ability to learn from practice.

### **Concerns about performance assessments for teaching**

Despite how positive outcomes of performance-based assessment in teaching and teacher education are getting widely recognized, there is still skepticism that this type of assessment can be a valid measure of teaching. One noted limitation is related to the content representation of performance-based assessments. These assessments are able to sample actual knowledge, skills, and dispositions of teachers directly, yet they are only able to ask a small number of questions or a small number of performances. This raises doubt about the ability of performance assessments to capture and measure the broader knowledge and performance domain expected in teaching.

Another skepticism is related to the subjectivity of the collected evidence. Written reflection is a very popular method for collecting teacher's responses in performance-based assessments. The best way to score the written reflection is for them to be read and scored by trained scorers. Often scoring rubrics and standards are provided for scorers to maintain scoring consistency. Still the scoring process raises questions because it is impossible to eliminate human being's subjectivity based in their personal, educational,

and professional background during the scoring process. This raises the issue of fairness as a deep consideration for performance assessments and how they are scored.

Additionally, there is a concern about a lack of evidence to prove positive outcomes of performance-based assessments on a large-scale in education settings. Positive outcomes can be understood very differently depending on the purposes of performance-based assessments and the context in which they are administered. Several studies demonstrate the positive outcomes of performance-based assessments on a small scale (Moss, 1992). There remains uncertainty about how performance assessments can maintain high quality of administration when using them on a large-scale or for decision-making purposes.

Empirical studies of performance-based assessments in teaching and teacher education research have addressed some of these concerns through validity arguments. The next section reviews and summarizes a selection of empirical validity studies.

### **Establishing Validity of Performance-based Assessments**

Performance-based assessments in teaching and teacher education have been developed and used in different ways. Some assessments have been developed by professional organizations and contracted testing organizations for administration (e.g., National Board for Professional Teaching Standards assessments administered by the professional organization and Praxis III administered by the Educational Testing Service). Local school districts or states have also provided leadership to develop performance-based assessments to evaluate their teachers. Also, higher education faculties have developed their own assessments for teacher candidates in their programs



(e.g., Performance Assessment for California Teachers--PACT). These Performance-based assessments are products of tremendous efforts of researchers and educators in teaching and teacher education. I have intentionally chosen a few of these large-scale performance-based assessments of teaching and teacher education and have reviewed their validity studies.

Many performance-based assessments in teaching and teacher education have research studies drawing on psychometric principles. In particular, performance-based assessments reviewed here carried out a series of studies to collect cumulative evidence to document the validity of the performance-based assessments. The most common validity studies examined content representation, construct alignment with external standards, scoring consistency, and fairness and bias studies.

### **Content representation**

Content validity studies evaluate content relevancy within the assessment by asking experts in the performance area to evaluate task relevance and content coverage with the authentic work within the performance field. The representation of tasks within performance-based assessments is linked to the underlying content of the assessments. As performance-based assessments have a limited number of questions or tasks they can ask of the test-taker, they should be developed well to measure what is intended. For example, the National Board for Professional Teaching Standards recruited 19 expert panels to evaluate the content validity of assessment exercises and scoring rubrics. The panels were experienced middle school teachers who were teaching the same content domains in National Boards assessments. At least 15 panels rated the exercises and

rubrics as important or relevant to identify qualified teachers in the domains (Benson & Impara in Jaeger, 1996).

The Interstate Teacher Assessment and Support Consortium (InTASC) carefully documented a process of development of InTASC standards. It recorded credentials and experience of contributing teachers and experts, charted the progress of selection of the principals and its components, and created a conceptual map illustrating links between the InTASC general principles and the content-specific principles (Moss, 1998). Another approach to evaluate the content representation for the InTASC was to interview the test-takers who completed portfolios based on InTASC principles (Collins, Schutz, & Moss, 1997). Among 18 recruited teacher candidates, 12 reported their teaching activities were reflected by the portfolio. Yet, only 4 felt that they had sufficient supports from teacher education programs that prepared them to complete the portfolio.

### **Construct alignment with external standards**

Several studies have investigated the alignment between an assessment's conceptual framework with local, state, and national teaching standards (Wei & Pechione, 2010). For example, PACT examined the alignment the Teaching Event tasks to the California Teacher Performance Expectations (TPEs) (Chung, 2008). Also, other studies examined their rubrics and expectations with external teaching standards including National Board standards and Danielson's Framework for Teaching (Porter, Youngs, & Odden, 2001). These investigations demonstrated how each assessment aligned with established standards and, thus, met requirements for accreditation and state teaching licensure.

**Scoring reliability**

Inter-rater reliability is one technique for measuring scoring consistency of multiple assessors. For example, PACT paired two scores of a sample of teaching portfolios completed by teacher candidates in California. The sampled portfolios were scored by a local faculty at the same university and one external scorer. This process was also carried out across all content areas. A consistency across pairs of scores calculated how paired scores were matched or not. The results showed that 91% of score pairs were exact matches or within one point (Pechione & Chung, 2006). Also, Praxis III conducted a study to examine how often two assessors agreed on their ratings on each rubric (Livingston, 1993). In this study, 43 different assessors evaluated 37 different beginning teachers' portfolios by using a scale of 1.0 to 3.5. The average difference between the assessors' ratings was 0.23. All paired assessors' ratings were within one point.

**Fairness and bias studies**

Validity studies also address issues related to fairness and/or bias that may occur in administration and scoring process of performance-based assessment. The National Board for Professional Teaching Standards investigated whether assessment tasks produced any gender or racial discrimination. One study found that there was evidence that white teacher candidates performed slightly better than black teacher candidates (Bond, 1997). There was no clear explanation for this gap between racial groups in their assessment performance, but this study helped test developers recognize issues about fairness and bias.

## **Underlying Validity Principals in Empirical Validity Studies**

The reviewed validity studies were built on two perspectives of validity theory: a traditional view of distinct types of validity (APA, AERA, & NCME, 1984) and Messick's unified view of construct validity (APA, AERA, & NCME, 1999, 2014; Messick, 1995). These two perspectives share a common goal of validity theory—the aim to offer practical guidelines to test developers and users for establishing strong validity arguments for assessments. However, those perspectives within validity theory have been developed with special interests and emphases.

### **Traditional view of validity**

The most frequently used perspective in the reviewed studies is a traditional view of distinct types of validity. This perspective separates validity into different types. For example, content validity, face validity, criterion validity, and construct validity were used to evaluate performance-based assessments of teachers.

Content validity is the estimate of how well a test represents the body of knowledge and actual skills that is being tested. This type of validity is associated with content representation and sampled questions of tests. Many educators whose main interests lie in measuring achievement tend to prioritize analysis of sampled test items, for example, alignment of test items against curriculum standards, to ensure the representation of curriculum content on the test.

Criterion validity examines how well a test predicts outcomes for another test. High correlation of scores between two tests is important evidence to investigate criterion validity. This type of validity is categorized into two sub-types; predictive and concurrent

validity. Predictive validity is associated with aptitude tests that predict an individual's future performance based on current knowledge and skills. For example, SAT scores are often used to predict the college freshmen Grade Point Average (GPA) by examining correlations between SAT test scores and the freshmen GPA. Concurrent validity is used to examine how well a new test is developed by comparing results to a well-established test. For example, the results of a teacher developed test in math can be compare with the SAT math scores. If a correlation between the two tests is high, it suggests that the teacher developed test is a valid measure of mathematical learning.

Construct validity infers an individual's possession of some abstract or hypothetical trait or quality (for example, intelligence or creativity). Among the three types of validity, construct validity was viewed as an "indirect" method of validation to be used when neither content nor criterion validity could indicate the degree to which a test measures what it was intended to measure. This type of validity is broader than the other two types, and it involves a combination of analysis of sampled test items and empirical evidence of correlation. Construct validity is explained in the next section with Messick's definition of unified view of construct validity.

This perspective of having distinct validity categories is helpful for making clear and practical guidelines for assessment development. Also, this perspective has contributed to the expansion of validity categories. However, this approach leads to a misunderstanding that investigations of one or selective types of validity studies guarantee validity of an assessment. This selective approach to validity studies may be too simplified to draw conclusions that an assessment is valid or not (Newton & Shaw,

2014). To avoid a false validation process, validity theorists reemphasize that we carefully understand the core definition of validity for accurate validation practices.

### **Messick's unified view of construct validity**

Messick's perspective to evaluate validity of assessments states, "validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessments" (Messick, 1995b, p. 1). He highlighted that validity is not a property of tests or assessments, but that the meaning of test scores must be interpreted and the meaning of the scores inferred within a framework of validity. He also articulated a general principle of evidence that should be examined to understand the meaning of test scores holistically (Messick, 1989).

The general principle includes six sources of evidence that are applicable to all assessments, including performance-based assessments (Messick, 1989, 1995a). First, the content aspect of construct validity examines the content's relevance, representativeness, and technical quality of the assessment. Second, the substantive aspect refers to the theoretical rationale for observing consistencies in assessment responses. Third, the structural aspect concerns the fidelity of the assessor's scoring procedures and scoring processes during the assessment. Fourth, the generalizability aspect examines how score properties and interpretations generalize to and across population groups, settings, and tasks. Fifth, the external aspect refers to the relationship between scores obtained in the assessment and scores obtained in other assessments used to measure the same construct.

Sixth, the consequential aspect is related to the consequences of the assessment for the person being assessed, including both intended and unintended consequences.

Messick's unified view is significant in that it integrates theoretical rationales and empirical evidence and sustains the claim that the test measures what it is supposed to measure, and therefore that it could be used for its intended purpose. Multiple lines of evidence allow for investigating the meaning of test scores more fully and completely.

Furthermore, Messick (1989) included social values and their consequences at the heart of validity theory. He stressed the importance of investigating consequences to understanding the meaning of test scores. This argument contributes to awareness that scores from a particular test can be interpreted and used in multiple ways, depending on individual responses and the contexts of assessment (Moss, Girard, & Haniford, 2006). Yet, it has been controversial among several validity theorists.

Overall, Messick provided important insights into multiple lines of evidence and expanded the scope of how we interpret meanings of test scores. However, this perspective heavily relies on technical considerations for establishing validity which sometimes assumes assessment designs that are test-item based. Some authors argue that alternative perspectives should be used to establish the validity of less standardized assessments such as performance-based assessments.

### **Alternative approaches to examining validity of performance-based assessment**

The two perspectives on assessment validity that have been discussed so far have a clear scope to investigate how well assessments are developed in terms of content representation, scoring process, and fairness issues. However, these perspectives are

limited in how they consider the actual impacts or learning consequences of performance-based assessments as evidence of validity. Some researchers have suggested alternative approaches to evaluate the validity of performance-based assessments. Their suggestions stem from understanding the unique characteristics of performance-based assessments especially with regard to the educative aspects of learning while engaging in the assessment itself.

### **Moss's situated inquiry for establishing validity**

With special attention to characteristics of performance-based assessments, Moss (1996) raised a question about the idea that all assessments should be evaluated by the same general principle (Messick, 1995a). She argued that this idea originally stemmed from an assumption of standardized assessments with standardized administration procedures and scoring systems in place. Under this assumption, contextual variances become evidence that theoretically decrease the validity and reliability of performance-based assessments (Shepard, 1993). However, when we think about authentic characteristics of performance-based assessment which are situated in dynamic and fluid contexts, Messick's general principle is not appropriate to evaluate the validity of performance-based assessments. It is important to remind ourselves that the dynamic and complex nature of the assessment is a key construct for providing authentic opportunities to learn by carrying out tasks in performance-based assessments (Darling-Hammond, Wise, & Klein, 1999).

Wise (2011) argued that the issues related to generalization and fairness and low score reliability are problematic. However, he suggested that these issues could be



addressed through stronger construct validity for measuring targeted traits or performance that we are not able to assess using standardized paper and pencil assessments. He also proposed to distribute assessment opportunities through a school year. This may help to partially compromise the low reliability of a single performance-based assessment by taking a composite over multiple assessment tasks and occasions. This innovative approach to validity echoed Moss's flexible standpoint to evolve validity theory to embrace distinctive features of performance-based assessments.

Moss, Girard, and Haniford (2006) captured a context where testing scores are always interpreted and used in different ways. Even though all assessments are validated and administered in systematic ways, the meaning of test scores can be different depending on moments of learning. They argued that both the test and the contexts of testing can be transformed by multiple factors. For example, same test scores are interpreted differently between teachers and other stakeholders. As teachers have more direct and sustain contacts with their students, teachers tend to interpret test scores by comparing students' growths over time in classroom. Other stakeholders including administrators and principals may have less information about students' performance in classroom so that they tend to focus on numbers to interpret test scores. As test scores can be interpreted in different ways depending on stakeholders' interests and positions, it is necessary to consider why we administer tests and collect test scores. Rather than focus on assessment procedures to evaluate performance-based assessment, it makes more sense to investigate the assessment interactions for formative or on-going learning within the contexts.

This perspective intends to be more practical and realistic about the approach to evaluating the validity of performance-based assessments. Also, another intention is to emphasize the roles of the test users (Moss et al., 2006). Moss and her colleagues challenge the viewpoint of validity which is dominant in empirical validity studies grounded on a traditional view of validity and Messick's unified view of construct validity. However, the suggested perspective does not intend to ignore or discard the current perspectives on validity (Moss 1996; Moss et al., 2006). It is very critical to maintain high standards of quality of tests and assessments by reviewing content representation, scoring processes, and standards alignment. However, one of the missing pieces is an understanding of the contexts in which the assessment is administered, which, arguably, has a large impact on the learning outcomes and consequences of the assessment.

### **Multiple dimensions of performance-based assessments**

Multiple research studies suggest alternative criteria to evaluate validity of performance-based assessments with special attention given to learning as a main purpose of the assessment. For example, validity theorists (Brown, 2004; Dierick & Dochy, 2001; Frederiksen & Collins, 1989; Haertel, 1991; Hambleton, 1996; Linn, Baker, & Dunbar, 1991; McDowell, 1995; Uhlenbeck, Verloop, & Beijaard, 2002) proposed different sets of validity criteria for performance-based assessment. In this section, I summarize and describe six types of validity criteria from their work.

By combining these criteria from various validity theorists, my intention here is to build a framework for examining the consequential validity of performance assessment

adoption. This framework serves both as an analytic frame for this study and as my effort to further define and synthesize what is meant by consequential validity within the measurement and assessment field. These criteria are educational consequences, meaningfulness, directness, transparency, fairness, and usability.

**Educational Consequences.** This dimension concerns the effect of performance-based assessments on learning and instructional improvement (Linn et al., 1991). A collection of evidence is needed to investigate the intended and unintended, positive and negative, and actual and potential effects of the assessments on how learners and teachers view the goals of education and adjust their learning and teaching activities accordingly. This dimension is originally developed from a concept of consequential aspect of validity (Messick, 1995a). A major difference from Messick's perspective is a focus on positive outcomes such as learning consequences or outcomes. Messick's (1989, 1995a) emphasis was to minimize unintended and adverse consequences for successful assessment administrations and implementation. However, the primary focus of this criterion is to investigate how assessment tasks and practices impact students' learning and how to improve learning practices to achieve a goal.

**Meaningfulness.** This criterion suggests that performance-based assessments should create a significant value for all participants including teachers and learners (Linn et al., 1991). Performance-based assessments should allow students to deal with meaningful problems that provide worthwhile educational experiences. McDowell (1995) stressed that for students to perceive an assessment as meaningful, they need to make a link between the assessment task and their personal interests. Also, performance-based

assessments for teachers should allow teachers to support their high-quality teaching. The assessments should be able to create a space where learners and educators place value on their learning and teaching activities in performance-based assessments. Investigations of attitudes or reactions to assessments would be critical for establishing the assessment's meaningfulness.

**Directness.** This dimension indicates the degree to which teachers should use test results or scores for modifying instruction. Linn, Baker, and Dunbar (1991) argued that indirect measurement indicators may cause misguided instruction. For example, using multiple-choice questions about writing does not capture actual writing competency compared to actual writing samples. Alternatively, performance-based assessment can capture actual ability to apply the knowledge of classroom management into teaching practice in classroom. Thus, directness of the assessment speaks to how tightly linked the test results are to the learning experiences and context in which the assessment is administered. The greater the need for interpretation, the less direct the assessment is for purposes of feedback and improvement.

**Transparency.** Transparency as a validity criterion means that performance-based assessments should be explicit about what is expected of all participants (Frederiksen & Collins, 1989; Linn et al., 1991). Learners need to be clear about what is expected of them, what tasks are required, and how to adjust their learning processes accordingly. Similarly, teachers should know and understand the entire performance-based assessment to understand their role in teaching and monitoring students and to be able to support student learning. To test the extent of how well the test is understood by

students and teachers, Hambleton (1996) suggested comparing how learners and teachers evaluate and judge a sample work. This criterion is especially critical to evaluate performance-based assessment since this type of assessment is complex.

**Fairness.** To scale up performance-based assessment, it is critical to examine validity in terms of the related issue of fairness (Linn et al., 1991). All assessments should be able to provide the opportunity for the test-taker to demonstrate his or her abilities and maximize their potential (Brown, 2004). A definition of fairness does not mean to promise the same amount of time or standardized conditions across varying contexts. However, it should be linked to creating equal learning opportunities. Each student has different skills and ability, so they should receive appropriate supports to ensure opportunity for students' learning.

**Usability.** The last validity criterion of usability is described by Brown (2004) as efficiency, by Linn et al. (1991) as cost and efficiency, and by Uhlenbeck et al. (2002) as practicability. This criterion is especially important because of the complexity of the task components in performance assessments, as well as the procedures and cognitive requirements in performance-based assessments. Evidence needs to be found that the investments in time, effort, and resources are justified for the positive effects such as improvements in teaching and learning (Hambleton, 1996). With a sense of practical use of assessments, all participants should find assessment tasks manageable in their contexts (Brown, 2004).

The reviewed studies suggest similar or different criterion to evaluate performance-based assessments. Frederiksen and Collins (1989) proposed that

assessments have “systemic validity” if they encourage behaviors on the part of teachers and students that promote the learning of valuable skills and knowledge, and allow for issues of transparency and openness, that is access to the criteria for evaluating performance. The accumulation of evidence of interpretations of assessment results by teachers, students, administrators, and policymakers, as well as the actions they take as a consequence, should be undertaken for educational assessment programs. Overall, the reviewed criteria imply the creation of a system to make a strong connection between curriculum, learning experiences, and assessments.

Uses of performance-based assessments have become popular in all educational contexts including teaching and teacher education. Numerous studies have been completed to investigate the validity of performance-based assessments in teaching and teacher education, yet there are skeptics.

In general, the key constructs of performance-based assessments are to measure targeted traits or performance directly and to provide learning opportunities concurrently. So far, empirical validity studies rarely capture the second part of the construct. Messick (1989, 1995) proposed consequential aspect of validity very briefly. He also described “the social consequences of testing as an integral part of validity” (Messick, 1993), suggesting that it is not a stand-alone validity concept, but an important component of validity evaluation. Shepard (1997) summarized multiple perspectives of validity theory and concluded that “the incorporation of test consequences into validity investigations” (p. 5) can be valuable. However, the question remains as to what extent consequential

validity is an independent concept of validity in assessment development and implementation.

Grounded on the reviewed literature, I created a framework for consequential validity that would further address the questions of the role of contextual and test-takers consequences as an aspect of assessment administration and interpretation. The framework for consequential validity has six dimensions: educational consequences, meaningfulness, directness, transparency, fairness, and usability. This set of six dimensions will contribute to the exploration of the role of contextual and test-takers consequences. Chapter three will revisit a set of research questions which were proposed in chapter one and along with the research design and methods.

## CHAPTER 3

### METHODS

The literature reviewed in chapter two recognized the unique characteristics of performance-based assessments in teaching and teacher education and how they provide learning opportunities for teachers and teacher candidates. Additionally, it suggested an alternative validity perspective, specifically a consequential aspect of validity, to evaluate the unique characteristic of performance-based assessments. Nevertheless, the concept of a consequential aspect of validity needs to be developed further. A key purpose of this study is to explore the concept of a consequential aspect of validity for performance-based assessments by examining the case of the state-wide edTPA implementation in Minnesota. This study did not intend to compare the process of edTPA implementation between institutions in Minnesota. Instead, this research aimed to investigate the overall edTPA implementation process in the context of Minnesota and to test the concept of consequential validity during implementation in the state.

After a three-year pilot study period (2011-2014), the Minnesota Board of Teaching authorized the edTPA as a performance assessment that meets the state legislated requirements for a performance assessment for teacher candidates beginning in 2014-2015. Based on this decision, edTPA was fully implemented across all institutions in Minnesota (see Appendix A for a list of all institutions and those selected for this study). Currently, edTPA is not used to determine Minnesota teaching license approval for individual teacher candidates. However, teacher education program's continual renewal by the state is based partially on the performance of candidates on edTPA.



Faculty at each institution have access to supporting materials developed through the Stanford Center for Assessment, Learning, and Equity (SCALE). For example, edTPA handbooks provide a detailed description of all of the edTPA requirements; the *Making Good Choices* handbook provides candidates with guidance for how to best use the edTPA for educative purposes; *Thinking Behind the Rubrics* describes the rationale behind the rubric levels and the nuanced distinctions between the rubric levels; and an online community discussion group for teacher educators hosted by the American Association of Colleges for Teacher Education provides several forums for communicating and clarifying the edTPA process. However, each institution must make local decisions about how to adopt edTPA based on needs and structures in their institutions.

### **Research Questions**

In seeking to explore a concept of consequential validity, the following questions drove this study: (1) what constructs should be included in a framework that describes the consequential validity of an assessment? (2) how can a consequential validity framework help describe the implementation of a standardized complex performance assessment administered in local contexts?

The research methodology used to explore these questions is discussed in detail below, with a description of the research design, a description of the selection process of the institutions and participants, and a description of the data collection and data analysis methods.

## **Research Design**

Overall, this study incorporated qualitative and quantitative methods to collect different but complementary data. The idea of mixing qualitative and quantitative methods is advocated because it is expected to expand the scope of data and deepen insights in research studies (Greene & Caracelli, 1997, Sandelowski, 2000; Tashakkori & Teddlie, 1998). However, Morse (1996) warned researchers to avoid a misguided assumption about conducting mixed method research design. Some researchers intuitively decide to use mixed methods because they believe that qualitative data is incomplete without quantitative data. To develop mixed methods design, the critical consideration is not just combining two methods to support each other, but well-defined justifications should be grounded to combine methods.

In this mixed methods design, the justifications were to establish triangulation to achieve validation between data and to have complementarity to fully explain and clarify results of the analysis. In order to explore consequential evidence of edTPA, it was critical to examine contextual and circumstantial factors associated with consequential evidence. To focus in on the local contexts, I used a three-tiered structure for organizing my data collection: the classroom level (Level 1), the program level (Level 2), and the institutional level (Level 3). Level 1 comprises the instructional contexts where faculty teach pedagogical and content knowledge to teacher candidates. Level 2 includes the program curriculum and practical experiences within schools for teacher candidates. Program and clinical coordinators play a key role in this level. For Level 3, I sought evidence from a leader of the teacher education program within an institution such as department chairs and deans to learn about how the overall organization implemented the

edTPA and the consequences of implementation on larger organizational concerns such as budgeting, scheduling, and staffing.

This research design is best described as multilevel research (Creswell, Plano, Clark, Gutman, & Hanson, 2003). It allowed me to use multiple forms of data to triangulate information as well as multiple levels of the organization to develop my findings. I collected both qualitative and quantitative data across all three levels during the same timeframe. The combination of both qualitative and quantitative data provided rich sources of evidence to examine consequential aspects of edTPA implementation.

## Data Collection Methods

This study used three data collection strategies: group interviews with teacher educators and teacher candidates, individual interviews with teacher educators and teacher candidates, and a survey of teacher educators. Qualitative data includes interviews and open-ended responses on a survey. Quantitative data includes 4-point Likert scale responses on a survey. I chose these data collection methods to focus on the participant perspectives with an aim to analyze positive and negative effects of the edTPA implementation in Minnesota and to further inform my understanding of the consequential aspect of validity in performance assessment administration.

Both focus group discussions and individual interviews with teacher candidates and teacher educators allowed them to explain their experiences and perspectives using detailed and specific examples. The data from the survey were used to check consistency of the perspectives of teacher educators at multiple teacher education institutions and compare them with the perspectives of teacher candidates. Table 3.1 shows data collection activities including data collection tools and the timeline of data collection activities.

Table 3.1

### *Timeline of data collection activities*

Time Period	Activity	
May, 2015	Send out emails to obtain contact information of potential participants	Survey development
June, 2015	Teacher candidate focus group interviews	
July, 2015		
August, 2015		

September, 2015		
October, 2015	Teacher educator focus group interviews	
November, 2015	Individual interviews with teacher candidates	
December, 2015	Individual interviews with teacher educators	Open survey
April, 2016		Close survey

### **Institution selection**

As this study aimed to investigate a statewide adoption of edTPA, it was important to choose institutions that would represent all institutions in Minnesota. When I set up criteria for institution selection, I used my observations and communications with faculty and staff during state edTPA meetings and the national meetings for sharing edTPA implementation strategies. I attended the Minnesota edTPA summit in 2014 and 2015, which provided opportunities to meet faculty and other staff. The meetings were great opportunities to get a sense of how different institutions have taken different approaches and were at different stages in their adoption of edTPA. For example, a faculty member mentioned that their institution just adopted edTPA in Fall 2015 even though the Board of Teaching had encouraged programs to participate in the pilot adoption the previous few years. Additionally, the faculty member explained a few reasons for late adoption decisions.

Consequently 15 institutions were selected with two criteria (see Appendix A for a list of all institutions and those selected for this study). First, the selected institutions represented the three types of higher education organizations within Minnesota – the University of Minnesota system, the Minnesota State Colleges and Universities

(MnSCU) system, and the privately run colleges and universities. I chose 4 institutions from the University of Minnesota system, 4 institutions from the MnSCU system, and 8 private institutions in Minnesota. Each type of institution was expected to take different approaches based on their resources, time, and other associated reasons. Second, the selected institutions had been active in piloting edTPA prior to the state-wide adoptions in 2014. With this criteria, I assumed that early adoption of edTPA would create more learning about implementation and better inform the consequential aspect of validity under examination in this study.

### **Participant selection**

I used a snowball or chain-sampling strategy (Miles & Huberman, 1994, 2002); Patton, 2002) to identify and invite teacher candidates and teacher educators in this study. This sampling strategy is a non-random or probability sampling technique that involves asking informants who one already knows to recommend or recruit future subjects from among their connections. This sampling strategy is often used in hidden populations which are difficult for researchers to access (Patton, 2002). Even though potential participants in this study were not truly a hidden population, this strategy was selected to access individuals who participated in edTPA implementation seriously.

To recruit participants for teacher candidate focus groups, I initially contacted faculty and program coordinators who have worked closely with teacher candidates in their program areas at the University of Minnesota, Twin Cities. I asked them to nominate 2-3 teacher candidates who completed their teacher preparation programs and submitted edTPA portfolios in Spring 2015. When I asked them to nominate teacher candidates, I clarified that a high score was not a critical criterion to nominate teacher

candidates for their participation in the focus group. My criteria were to interview teacher candidates who had participated fully in the edTPA process. The criteria for selection of participants were as follows:

Ideal sample of teacher candidates

- 24 teacher candidates at the University of Minnesota-Twin Cities campus, eight representing early childhood and elementary education programs, eight representing secondary education, such as mathematics, social studies, science, and literacy education programs; and the final eight representing K-12 education programs from music, art, physical and health, and English Second Language (ESL) education programs.
- All teacher candidates completed their edTPA requirements in Spring 2015.

Altogether 22 teacher candidates participated in this study. Table 3.2 shows the participants' program background. All program areas were categorized into three groups depending on what grade level teacher candidates taught.

Table 3.2

*Teacher candidate participants and their University of Minnesota licensure program*

	Licensure Program	Number of teacher candidates
Primary	Early Childhood	2
	Elementary	4
Secondary	English	3
	Mathematics	2
	Science	2
	Social Studies	1
K-12	Music	1
	Second Languages	6
	Physical & Health Education	1
	Special Education	1
Total		22

To identify teacher educator participants for the study survey, I contacted core faculty who have been involved in edTPA implementation in their institutions and have participated in the state implementation conferences I have attended. I asked them to identify at least one teaching methods instructor, one program or clinical coordinator, and one program leader in their institution who they thought would provide informed responses for my survey.

Ideal sample of teacher educators and institutional leaders

- 65 teacher educators and institutional leaders from 16 institutions with teacher education programs in Minnesota.
- 20 of these 65 participants from the University of Minnesota-Twin Cities campus asked to complete a survey and participate in a focus group interview.
- 45 of these participants from 15 institutions that have worked on the edTPA implementation process for at least two years in Minnesota and asked to complete the survey.
- Teacher educators in this study were categorized into three different groups. The first group includes methods or foundations course instructors who have designed coursework to support teacher candidates for preparing and completing the edTPA. The second group includes program and clinical coordinators who decide curriculum changes in a program areas or changes to practicum requirements. The third group includes the dean, department chairs, or teacher education directors in the teacher education program. I anticipated that each of these groups would provide similar and different perspectives on



the implementation of the edTPA and its impacts on each teacher education program and the teacher candidates.

Overall, 11 teacher educators at the University of Minnesota participated in this study.

Table 3.3 indicates the participants' program background.

Table 3.3

*Teacher educator participants at the University of Minnesota and their licensure program.*

	Licensure Program	Number of teacher educator
Primary	Elementary	2
Secondary	English	1
	Mathematics	1
	Social Studies	1
K-12	Agriculture	1
	Arts	1
	Music	1
	Second Languages	2
	Special Education	1
Total		11

**Interviews with teacher candidates.** Three one-hour focus group interviews were held with 6 teacher candidates in June 2016 and 16 individual interviews were conducted in November 2015 (See Appendix B: Focus group interview protocol for teacher candidates). These interviews primarily addressed Level 1—the classroom and instructional context of teacher preparation—and helped me to understand teacher candidates' perceptions regarding the implementation of edTPA in their coursework and programs. The intention of these focus group interviews was to obtain teacher candidates' perspectives of the edTPA preparation in their programs and the general learning outcomes they perceived as being part of the edTPA.

**Interviews with teacher educators.** I conducted focus group discussions with 11 teacher educators at the University of Minnesota (See Appendix C: Focus group interview protocol for teacher educators). The purpose of these interviews was to gain insights about edTPA implementation at this university through the lens of faculty who were highly engaged with their programs.

Different from teacher candidates, teacher educators may have different perspectives based on their roles and responsibilities in their program and institution. These interviews were positioned in Level 1 to address areas of classroom interactions and instruction, Level 2 to address issues related to overall program design for edTPA implementation, and Level 3 to understand institutional decisions during the adoption of edTPA. The focus group interview protocols were prepared to support participants' thinking processes, but the questions were modified and rephrased depending on the responses during the focus groups.

**Survey across fifteen institutions.** As I described above, it was critical to select participating institutions which had enough experience with edTPA to fully address the range of consequential questions I identified in my framework. It was also important that the participants within the institutions had a broad enough understanding of how their institution operates that they understood the contexts in which the survey questions may manifest locally.

I administered a survey to teacher educators across 15 institutions. The selected institutions represented three different types of higher education organizations in Minnesota. Faculty who have implemented edTPA in each institution were invited to complete a survey. This group included method-course instructors, program

coordinators, clinical coordinators, department chairs, and deans. The survey included 69 questions with items measured on a 4-point Likert scale and 6 open-ended questions. This survey was released in early December 2015, and was closed in late April 2016. It took approximately 30 minutes to complete the survey. All responses were automatically saved on the Qualtrics platform.

As I closed the survey in April 2016, 33 teacher educators had started the survey and 22 of them had completed the survey, resulting in a response rate of 34% of the total sample of 65. Table 3.4 provides institutional backgrounds of the survey completers.

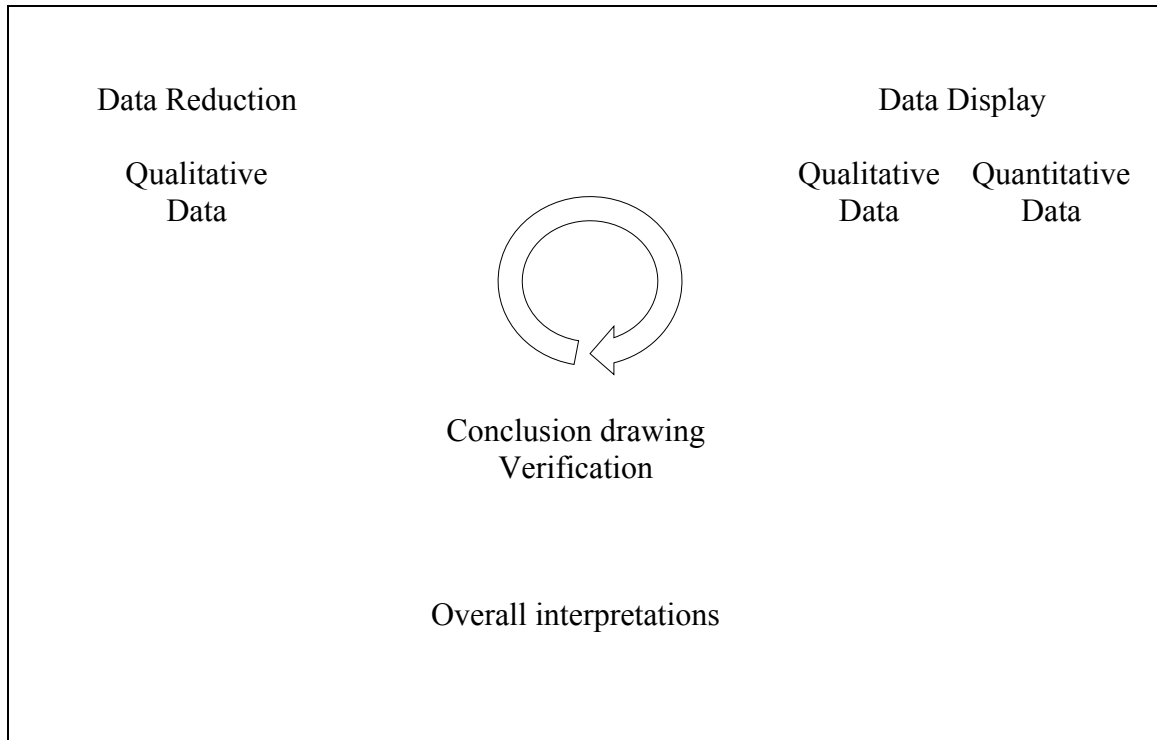
Table 3.4

*Survey completers and their institutional background*

Institution type	Number of teacher educators
University of Minnesota system	5
Minnesota State Colleges and Universities	7
Private Colleges and Universities	10

### **Data Analysis Methods**

Overall, I used Miles and Huberman's (1994) concurrent data analysis comprising three components: data reduction, data display, and conclusion drawing, in an on-going process. Furthermore, I modified a data analysis model from Miles and Huberman's (1994) and the modified model is showed in figure 1. Quantitative data emerges with qualitative data during data display.



*Figure 1.* Modified Data analysis model from Miles and Huberman's (1994)

Data reduction began by analyzing transcriptions of all qualitative data from interviews and open-ended survey responses. “Data reduction is a form of analysis that sharpens, sorts, focuses, and organizes data in such a way that ‘final’ conclusions can be drawn and verified” (Miles & Huberman, 1994, p.11). During this process, the primary task was to define key themes and to develop an analysis plan (Namey, Guest, Thairu, & Johnson, 2008). By using the six dimensions of the consequential validity framework (educational consequences, meaningfulness, transparency, directness, fairness, and usability), I extracted direct quotes from interview data and open-ended survey responses and coded for each dimension. The data from two groups, teacher candidates and teacher educators, were treated separately. For teacher educators, I analyzed each dimension across the three different levels of classroom, program, and institution. The extracted direct quotes were coded into either positive or negative experiences and perceptions

related to edTPA. By reviewing the extracted multiple quotes under each dimension, a few themes were identified for summarizing and displaying patterns of the data. This was called “first level coding” (Miles & Huberman, 1994, p. 56).

The first level coding resulted in multiple data fragments coded to the six dimensions of the consequential validity framework based on conceptual and word alignment of the data and the framework dimension. For example, teacher candidates described how their faculty members and program coordinators introduced edTPA to them. These descriptions of introduction activities were coded under the dimension of transparency as they represented moments of helping the teacher candidate learn about the assessment process, thus making the assessment clearly understood by all participants. As introductions to teacher candidates were critical activities to understand the assessment clearly, these descriptions were coded into transparency. Also, both groups of participants directly used language of fairness during the interviews. So, discussions around fairness were coded in the fairness dimension.

To develop a more systematic data display, I created a matrix table organized around the dimensions of the framework and each of the two groups of participants. I looked for patterns within the data fragments assigned to each of the dimensions and labeled the fragments with themes. As I completed the matrix, I could identify agreement or disagreement among participant groups around similar themes and identify connections across the six dimensions.

For the survey data, I calculated descriptive statistics for each of the survey items and created displays for each framework element. A rationale to use 4-point Likert scale was to force potential respondents for capturing the best way to describe their experience

and perception. To gain clear pattern of the survey data, I simplified the scale into two categories: agree and disagree.

The merging of two data sources served to draw conclusions about how edTPA has been adopted and perceived by multiple participants. The proposed consequential validity framework served to investigate local contexts of a performance-based assessment systematically and to examine links between the local contexts and educational consequences. Also, this process allowed me to examine the proposed conceptual framework for proposing practical guidelines for validating consequential aspect of performance-based assessments.

Chapter four reports detailed findings from the data collected in this study. Contextual information across three levels within institutions and personal perceptions and experience were valuable data to draw full understanding of assessment implementation and its consequences.

## CHAPTER 4

### RESEARCH FINDINGS

#### **Analysis of Consequential Validity in the Implementation of edTPA in Minnesota**

This study employs six dimensions of a proposed consequential validity framework which was discussed in Chapter two. *Educational consequences* refer to how assessment tasks and activities have impact on students' learning and how to improve learning practices to achieve a goal. *Meaningfulness* refers to how participants perceive the underlying value of engaging in the assessment activities and processes. *Directness* speaks to how the results can be immediately interpreted by the assessor and the participants with little need for technical translation or indirect interpretation. *Transparency* means that the assessment and its administrators should be explicit about what is expected of all participants. *Fairness* promises to provide equal opportunities for all students to demonstrate their abilities and knowledge. And *usability* refers to the practicality of administration of the assessment regarding investments of time, effort, resources, and management by all participants.

In this chapter, I use the framework with six dimensions to address one of the research questions posed in this study: how can a consequential validity framework help describe the implementation of a standardized complex performance assessment administered in local contexts? I systematically investigated the recent adoption of the edTPA in Minnesota, including contexts, processes, and consequences of the adoption of this performance-based assessment using this framework.

edTPA has been implemented in Minnesota since 2011-2012, yet little research is available to gain an understanding of how this assessment is adopted and used in local

contexts. It is critical to investigate how an adoption of a new assessment has created impacts on teacher candidates and teacher educators. Currently, edTPA is not used to make teacher licensure decisions. Understanding these consequences may contribute to determining the potential use of this assessment for teacher licensure decisions. This study also aims to enhance our understanding of the concept of consequential validity in educational assessment that can lead to practical suggestions for implementing complex performance assessments in local contexts.

This chapter is organized by groups of participants who were surveyed and interviewed in this study. First, analysis of data from teacher candidates presents a critical examination of the first-hand experiences with the edTPA, providing insights from the classroom level of implementation. Next, data from teacher educators provide details and visions from the perspective of faculty, and provide insights from the classroom as well as the programs and institutional level of implementation.

### **Teacher Candidates' Perspectives**

The interview questions asked teacher candidates how they experienced edTPA as a part of their teacher education preparation. Throughout the interviews, teacher candidates described details of assessment implementation in the classroom and program. Also, they reported how different implementation approaches created unique learning opportunities with positive and negative consequences. The consequential validity framework allowed me to explore multiple aspects of the assessment implementation based on teacher candidates' perspective.

**Educational Consequences.** Teacher candidates indicated both positive and negative educational consequences of edTPA during interviews. Positive consequences were



associated with impact on their learning process and learning outcomes, aligned with the expectations underlying edTPA design. Consequences were viewed as negative when they were associated with unexpected outcomes such as when learning was hindered or interrupted.

**Positive Consequences.** Interviewees reported that several aspects of edTPA served to improve teaching practices in their classrooms. First, reflection practice was the most beneficial to candidates. The reflection process stimulated them to consciously consider the needs of students' learning. Concurrently, the process made them think about how to adjust and improve their teaching practices to meet the needs of students' learning. One interviewee said:

...it helps you to really focus on what you've taught, what helps students learn what you teach, and it helps you focus on how what you taught either benefited or not the students that you're teaching. And on the data you collected, you could really drive your instruction going forward rather than moving onto the next things as planned.

The reflection process generated impactful thinking processes to make a tie between learning of students and learning of teachers.

According to interviewees, the reflection process stimulated them to acknowledge some aspects of teaching they had not considered and reflected on deeply prior to their edTPA experience. One teacher candidate commented, "I think it was kind of above and beyond anything I would ever do in the real world... I liked the way it made me reflect." Interviewees asserted that the task *Context for Learning Information* encouraged them to focus on students' learning. This task asked teacher candidates to collect information

about students' personal and educational background and needs before they actually developed lesson plans and instruction. Some teacher candidates volunteered to review their portfolios before the interviews in this study. According to them, they recognized that the reflection practices actually supported them to focus on students' active involvement during their teaching. According to teacher candidates, edTPA required them to plan out lessons with a full searching and understanding of students' background and needs, changing their perception of teaching from what teachers can do to what students need to learn.

Also, teacher candidates reported that edTPA as a self-assessment was powerful to examine behaviors of teaching by using videos. Videos captured vivid moments of teaching and allowed the candidates to evaluate their verbal and non-verbal behaviors while teaching. A teacher candidate stated:

...as a beginning teacher, we don't notice a lot of what we do in the classroom, or how our body is, or what we say exactly...so I think that, you know, doing a recording like that really does help you see, how do I use my body? or, if I'm asking students to be calm but I'm moving around, how does that affect them?

As teacher candidates reviewed their own teaching moments, it encouraged them to see how they physically and linguistically behaved unconsciously.

The focus on academic language created another learning process for teacher candidates that was a positive consequence. Not all candidates felt confident in how to integrate academic language during their lessons. However, teacher candidates who intentionally used academic language concepts in lesson planning asserted that it challenged them to evaluate their word selection during lessons. Also, it supported

teacher candidates to observe students' language behaviors in the classroom and to make a plan to transition from social language to academic language. For example, students described a cylinder as a *circle thing*. A candidate recognized that some students could not recall the term of cylinder correctly. So, she created a quick activity to practice the term before moving on to the next planned activity.

**Negative consequences.** Unexpected aspects of the edTPA assessment process that minimized or diminished the learning process of edTPA included the burden of reflective writing and the feeling of working on edTPA as only a means of working to the test requirements. Teacher candidates reported that they sometimes felt like the writing requirements created a burden and distracted them from learning. One teacher candidate expressed her feeling, "I...got to the point where the reflection piece stopped being useful and it started being just more like a regurgitation a little bit." Other teacher candidates reacted similarly to this issue. For example, one teacher candidate decided not to participate in school events and conferences to save time for studying edTPA handbooks and writing reflections. Four teacher candidates reported that it took away opportunities to build relationships with students, parents, cooperating teachers, and other staff in the school. As teacher candidates dominantly viewed that flawless writing was expected in the assessment, their perspective on edTPA was that it was a distraction from what they perceived to be the real work of teaching.

In some cases, teacher candidates recorded instructional videos early in their edTPA process and then waited until much later to write the commentaries and reflection just before the submission deadlines. Even though the edTPA directions suggest that teacher candidates take time to review and reflect on their teaching practices before they

move on to the next lesson, this did not happen for all of them. For candidates who completed the activities without linking the activities to the daily practice of reflection and improvement, the edTPA became a set of disconnected activities that they had to complete in order to complete the assessment. Rather than

**Meaningfulness.** This dimension investigated how teacher candidates perceived edTPA and its related activities as a valuable opportunity for learning how to teach. They confirmed that the underlying concept of edTPA was a valuable framework to think about and assess teaching effectiveness and to support teaching improvement. The tasks and candidate responses to given tasks seemed to support candidates' motivation and engagement with edTPA. In addition, candidates identified external motivations such as accountability and faculty expectations that encouraged the candidates to take the assessment activities seriously.

In the interview responses, the candidates' placed different values on the edTPA at different times and occasions. Teacher candidates across program areas identified valuable experiences of edTPA from different sources. They viewed that the assessment design itself and interactions with faculty members created meaningful experience during assessment activities. One teacher candidate said:

...about really being aware, being planful about trying to include opportunities for students to see themselves reflected in a lesson. That was meaningful because it did spur thinking beyond the edTPA about that and something I still carry forward in my practice.

This teacher candidate recognized that the tasks of edTPA were a valuable framework to think about and assess teaching progress and to make a link between their teaching and

students' learning in the classroom. Also, this candidate viewed the meaningfulness of edTPA as not only a way to evaluate teaching practice as a teacher candidate, but extended its value to what it might offer in their first year of teaching. Another candidate confirmed this:

it is an okay assessment to measure my current teaching ability, however it is hard to tell my ability as a novice teacher. This assessment will be a great resource to check my growth a few years after.

One third of interviewed candidates wished to use this type of assessment during induction with their mentors. In some cases, they didn't recognize the values of the assessment during student teaching because heavy workloads and other responsibilities consumed them during their program. Upon reflection on the process during the research interviews, two candidates said that conversations during the interviews supported them to reflect on the overall process of this assessment, its values, and their accomplishments documented in the assessment activities.

The most difficulty in engaging candidates with the assessment was that not all teacher candidates were motivated to participate seriously. Several reported that they were under pressure with other tasks and responsibilities and that there was not a strong external stimulus that was pushing them to engage deeply with the assessment processes. Without accountability, candidates reported that they put the edTPA requirement as their lowest priority during student teaching. One candidate said:

[I]t doesn't affect any of your grade. So that kind of took away the meaning behind it. We're told like, oh, it's not – you don't have to necessarily pass it, so no big pressure, but you have to take it. When they put an emphasis on the fact that it

doesn't affect our ability to get a job, it just seemed like it was overwhelming for no reason.

Motivation is a critical component in engaging in a performance-based assessment system. While some candidates seriously engaged with the assessment, several felt the absence of an external need to engage seriously and did not engage wholeheartedly. It is important to recognize that most teacher candidates were not motivated internally without supports from teacher preparation program. Without understanding values of the assessment, it was difficult for the teacher candidates to imagine authentic learning from participating in the edTPA assessment.

In addition to the external accountability for engaging in edTPA, teacher candidates also asserted that high expectations from faculty motivated them to put more effort into their performance on the edTPA tasks. One teacher candidate said, “We know they [faculty] care about it more, so we try harder.” Different from accountability, faculty’s expectations created gentle pressure and more local motivations to participate fully and try to put forward strong performances on all aspects of the assessment. Another teacher candidate reacted similarly:

She [an instructor] was very, very gung-ho about it. So I think, sort of, just that attitude kind of was bestowed upon us. She wanted us to do well. I think that was an implicit expectation of us.

Interviewees expressed a strong attachment with their faculty instructors. Depending on faculty’s portrayed attitudes and expectations to edTPA, candidates quickly made conclusions about how much effort they needed to put into their edTPA. Accordingly, if the teacher candidates received the message from their program faculty that the edTPA

did not matter, they internalized this message and did not put much effort into the completion of the assessment tasks.

In regard to gathered evidence, teacher candidates need some time to explore the details of this assessment. Without awareness or understanding of any values and benefits of the assessment, teacher candidates struggled to be motivated. Teacher candidates pointed out roles of teacher educators in helping them understand intended goals and uses of this assessment and reported that the teacher educators' expectations of their engagement with the assessment played a significant role. The role of the teacher educator was not to teach them how to take the test, but teacher educators supported teacher candidates in engaging in the edTPA as a process of gathering evidence of learning to inform education-related decisions.

**Transparency.** According to the teacher candidates, it was vital to have a clear understanding of the assessment activities and process, including assigned tasks, rubrics, expectations, and submission deadlines for submitting edTPA successfully. Three main themes are discussed in this dimension of the framework: support systems for teacher candidates, workload while completing edTPA, and writing anxiety.

First, while each program took different approaches to introduce and administer edTPA, common strategies across program areas were reported by teacher candidates. Some strategies identified in the teacher candidates' interviews include:

- Reviews and analysis of edTPA handbook and rubrics with faculty and supervisors
- Analyses of completed examples from previous year
- Practicing full or partial tasks in coursework before the actual edTPA

- Technology supports from the program, department, or college
- Optional weekly meetings to ask questions and to work on edTPA tasks with peers during student teaching

These examples supported teacher candidates to understand the required tasks, details of administration, expectations of both the local program and the external edTPA scoring process, and uses of the assessment results. The approaches varied in terms of time-point of introduction, leadership within a program, and the level to which edTPA was embedded into existing curriculum. Depending on program decisions about how and when edTPA was introduced, teacher candidates received different types and levels of supports. With understanding of the administration information, candidates felt ready to participate in the assessment.

Candidates reported that they faced the most difficulty when making decisions about what elements to be included and discussed in their lesson plans, reflections, and required supporting documents from practice (video and student work samples). The most helpful activities to support these decisions were conversations and feedback from faculty, supervisors, peers, and cooperating teachers with intentional guidelines from faculty.

When teacher candidates were asked about their first impression or reaction to edTPA, they typically responded that they were overwhelmed and felt lots of stress and pressure. Yet, they pointed out that the messaging about edTPA when faculty introduced the assessment was problematic. One teacher candidate commented:



it seemed really overwhelming at the beginning. But then when you started powering through it, it really wasn't that bad at all, ... I think it gets worked up to be really hard work that you won't be able to do.

This teacher candidate points out that they were led to believe that the assessment would be really difficult to complete and to pass and that this message was not helpful to them. While teacher educators' intention may have been to ready teacher candidates for the intense schedule during student teaching while completing edTPA, the message elevated negative emotions and threatened teacher candidates unintentionally.

Teacher candidates characterized edTPA as a comprehensive capstone project for them to articulate everything what they had learned in their programs, rather than as a process for learning during their program. The candidate quoted above suggested that if the faculty could pause and give quick tips during coursework, or talk about the edTPA related to what they were doing in coursework it would be helpful to teacher candidates to have a real sense of workload and make a plan for themselves.

Since edTPA heavily relies on reflective writing, teacher candidates automatically made an assumption that it would be easier to a particular group of teacher candidates. One teacher candidate said, "I am not a good writer...maybe edTPA is much easier for literacy people." While writing skills are not a criterion that is scored in edTPA, teacher candidates assumed that their writing ability impacted their official scores anyway. Another teacher candidate interpreted her experience in this way, "...if you can't put your reflections on paper in a precise manner, then the edTPA might not assess you appropriately..." Throughout interviews, I observed that teacher candidates misunderstood how their writing ability would be reflected in official scores. They

believed that writing skill was very critical to getting a high score on this assessment. Their beliefs caused apprehensive and negative feelings about writing.

**Directness.** This dimension investigated whether edTPA provided direct evidence of what teacher candidates know and can do as teachers in the classroom. The assessment design anticipated that each teacher candidate would receive context-specific and individualized feedback from teacher educators to improve their teaching. The interviewed teacher candidates reported that the collected evidence as part of edTPA supported them to review their teaching performance based on what they were actually doing in the classroom. Yet, there were barriers to getting actionable feedback from teacher educators that they could apply to their teaching and their edTPA materials. Teacher candidates reported that using video was a powerful way to review their behaviors during teaching. Videos captured vivid moments of teaching and helped teacher candidates self-assess their teaching in direct and immediate ways that influenced how they viewed their own teaching behaviors.

If programs fully embedded edTPA in existing coursework in advance of student teaching, teacher candidates gained opportunities to receive direct and actionable feedback from faculty members based on collected evidence on edTPA materials during the course instruction. However, if programs decided to wait until student teaching and administer a real edTPA only at that point, teacher candidates didn't get enough feedback from faculty members. One teacher candidate said:

She [an instructor] couldn't say anything ... because they're not allowed to actually help us with it...they give the feedback, it's not really clear. It's from

other graduates as well, because it isn't really clear how much they can give you, give teaching feedback to candidates.

Other teacher candidates perceived that it was an issue to get prompt feedback from faculty members while preparing materials to submit for the final edTPA. Similarly, three teacher candidates from different programs reflected that feedback from faculty members and supervisors was not directly helpful. This finding revealed a conflict between ethical coaching for the final edTPA submission and direct coaching on other aspects of edTPA prior to final submission.

**Fairness.** edTPA has been administered in a standardized manner using an identical assessment method, content, and scoring procedures. Yet, teacher candidates reported that contextual variations created different opportunities for different candidates to demonstrate their teaching knowledge and skills through the edTPA tasks. This dimension examined what aspect of implementing edTPA generated issues related to fairness from teacher candidates' perspectives including the support that they had and the barriers they faced.

Teacher candidates reported that cooperating teachers were significant to creating opportunities to reflect on their performance in edTPA. As cooperating teachers were familiar with this assessment, teacher candidates were supported to prepare edTPA without unexpected problems. One teacher candidate remembered the first meeting with a cooperating teacher. During this meeting, they quickly overviewed requirements and expectations with each other. Also, the cooperating teacher made a few suggestions for completing edTPA with consideration of a scheduled field trip, MCAs testing dates (state administered standardized testing), and other events in the class and the school. When the

time came to record lessons, the cooperating teacher explained to students why the student teacher needed to record lessons and how they could support their student teacher. Another cooperating teacher slightly changed a schedule on the day when a teacher candidate recorded their edTPA lessons so that the students and the teacher in the video could focus and record clear voices. The teacher candidates reflected that the most beneficial experience was to feel supported and welcomed so that they could be a teacher in others' classrooms.

Teacher candidates described a few settings when they felt a lack of opportunity to demonstrate their teaching ability. Among the interviewees, three teacher candidates were assigned to cooperating teachers who asked them to use lesson units developed by curriculum specialists in schools or private company. Teacher candidates perceived that it took away opportunities for them to demonstrate their ability to prepare a lesson sequence. One teacher candidate said, “depending on the school you're teaching at, you don't always have the freedom to choose the assessments you're using or what you're teaching. It is not fair.” With a clarification during the interview, this teacher candidate agreed that their perceived lack of freedom is same as feeling a lack of ownership as a student teacher.

Interview data suggested that contexts of student teaching led to different experience with edTPA. Assigned practicum sites and characteristics of cooperating teachers were not controllable. Still, the uncontrollable aspect in a performance assessment process caused disadvantages to a group of teacher candidates.

**Usability.** This dimension examined what aspects of the assessment process supported or hindered teacher candidates' participation in edTPA. Teacher candidates reported that

edTPA demanded a significant amount of time, effort, and energy. The biggest challenge was to find sufficient time to focus on edTPA. Teacher candidates reflected that it would be more manageable if they were able to estimate time and workloads along with other roles and responsibilities they would have while student teaching.

Teacher candidates who reported that they had an acceptable level of workload showed three common experiences. First, they had a comprehensive overview of edTPA and, second, they practiced all the tasks with feedback from faculty members in advance of student teaching. Finally, they had on-going conversations with faculty members and peers about tasks, rubrics, expectations and other aspects of the assessment. One teacher candidate said:

when it finally came to the real edTPA, I knew what to expect. And I knew how to plan for it within my student teaching. I knew how to pick what lesson to do, and what to focus on. I think that was a big help.

This teacher candidate received a comprehensive engagement with edTPA and gained a realistic sense of the edTPA process. Other groups of teacher candidates who independently reviewed and completed the tasks in edTPA spent tremendous time learning to understand this assessment without supports from faculty members and program cohort members. A teacher candidate in this group reported:

If our professors during our first semester were aware of what we were gonna be doing in the second semester, as we're going through plans, they could pause and say, "Hey, this is something you're going to be using – so, that TPA assignment – you're gonna have to turn in your lesson to that TPA." And to have us pull those

pieces out during the first semester so that – I think that would make the actual process of it easier and save our time.

Teacher candidates argued that alerts from faculty members were helpful. One teacher candidate stated:

Like even at the beginning of the semester [student teaching], we were talking about when you should think about doing it and all the different steps of that process; because they said, you know, "Don't do it the same time as your lead teaching; that'll be too much." That was a really big thing.

Once student teaching began, many teacher candidates tended to put aside edTPA until the end of student teaching and this created a lot of burden for them. Alternatively, for a few candidates who were interviewed, faculty kept monitoring students' progress of completing edTPA requirements and reminding them of the tasks, and teacher candidates were able to make a plan for data collections from the beginning of student teaching.

In terms of technical supports that increased the usability of the edTPA for candidates, programs and departments provided basic sessions on how to edit, save, and upload video files. While the technical supports made the assessment process more manageable, candidates also asserted that they desired more supports in how to analyze and make sense of their teaching after they completed their video capture. These kind of supports would have made the process more intellectually meaningful while also increasing the efficiency of how candidates spent their time during the analytic process.

### **Teacher Educators' Perspectives**

The survey for teacher educators asked how they experienced the edTPA implementation based on their perspectives in the classroom, at the program level, and at

the institutional level. For each level, a set of survey questions was asked to examine each of the six dimensions of the consequential validity framework during edTPA implementation. Additionally, this survey included open-ended questions to obtain more detailed information from the teacher educators' experiences. The survey data provides a more general picture of edTPA implementation across the state of Minnesota. Interview data from teacher educators at one institution provides a richer and deeper explanation of each dimension of the consequential validity framework. The interviewed teacher educators were asked to explain how edTPA created learning outcomes for teacher candidates and changes in instructional practices for teacher educators, if at all. Throughout the interviews, teacher educators shared what changes had been made to instruction and program design during edTPA implementation and why the changes happened. Each framework dimension illustrated how edTPA was implemented in classrooms, programs, and institutions with positive and negative consequences. Due to having too few participants who represented the institutional level (for example, deans and directors of teacher education), analysis within some dimensions combines both program and institutional level results.

**Educational Consequences.** This dimension investigated what aspect of edTPA created educative opportunities to understand learning goals in teacher education and to modify learning processes accordingly. Teacher educators identified both positive and negative educational consequences by implementing edTPA in their institutions.

**Classroom level.** Teacher educators predominantly viewed that edTPA supported candidates' learning to teach: 85.5% of teacher educators agreed that engaging with edTPA helped teacher candidates to develop their knowledge and skills for teaching

as shown in Table 1. Even further, 90% viewed that the interactive cycle of lesson planning, instruction, and assessment in edTPA helped their teacher candidates understand a holistic nature of teaching. Additionally, 61% of teacher educators viewed that edTPA helped their teacher candidates identify areas of instructional practice that they can improve. At the same time, this new assessment has added new learning opportunities for teacher candidates in courses from the teacher educators' perspective. The respondents had mixed reactions about whether the addition of edTPA to the teacher licensure curriculum interfered with candidates' learning experiences, with 54% reporting that it did interfere and 46% reporting that they didn't think that it interfered with learning.

Table 4.1

*Survey results in educational consequences at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
engaging with edTPA helps our teacher candidates to develop their knowledge and skills for teaching.	4.5% (1)	9.0% (2)	49.5% (11)	36.0% (8)
the interactive cycle of lesson planning, instruction, and assessment in edTPA helps our teacher candidates understand a holistic nature of teaching.	0.0% (0)	9.0% (2)	54.0% (12)	36.0% (8)
edTPA helps our teacher candidates identify areas of instructional practice that they can improve.	4.5% (1)	31.5% (7)	36.0% (8)	27.0% (6)
the added burden of edTPA interferes with our teacher candidates' learning experience	14.5% (3)	31.5% (7)	36.0% (8)	18.0% (4)
implementing edTPA has added new learning opportunities for our teacher candidates in courses.	9.0% (2)	9.0% (2)	63.0% (14)	18.0% (4)

*Note: Due to rounding, percentages do not appear to add up to 100%.*



All interviewed teacher educators argued that edTPA didn't reveal new learning topics to the current teacher preparation program except a concept of academic language. One survey respondent reflected that edTPA provided new learning opportunities for teacher candidates by introducing the concept of academic language, with one methods instructor reporting, "Well, I think that it has really been a learning tool for instructors. We've had conversations like, about academic language that we would have never had those conversations except for edTPA."

In addition, teacher educators reported in interviews that edTPA revealed how much candidates mastered the knowledge and skills covered in coursework and what areas they had difficulty applying the knowledge and skills in real classroom settings. They also reported that edTPA provided a scaffold for candidates to reflect deeply on why and how to teach.

One methods instructor observed that candidates experienced difficulty making decisions about how to change and modify their lesson plans during student teaching. By having conversations with candidates, this instructor realized that they were more familiar with thinking about what to teach rather than why to teach. As teacher candidates participated in edTPA, they had a structured process for reviewing their teaching by using the interactive model of planning, instruction, and assessment. This model supported teacher candidates to break down teaching into small components and to think cognitively about how each component needed to be connected for successful teaching. The instructor claimed that the model contributed to stimulating teacher candidates toward analytic thinking which was an effective practice for learning about teaching.

**Program and institutional level.** Generally, program faculty and institutional leaders viewed the adoption of edTPA as having positive educational consequences for program improvement. Teacher educators in the survey reported that the expectations and rubrics were helpful in revising learning content of coursework and in improving the learning experiences for teacher candidates. However, only 40.5% of them strongly agreed or agreed that edTPA was helpful to determine a teacher candidate' readiness to teach. One program leader reported that "it[edTPA] does force a little bit more of the interactive model and focus on individual student teaching. It's been a useful mechanism for that." She reflected that her program was more likely to focus on delivering lessons to the whole group rather than focus on individualized curriculum for the teacher candidates. The hesitance to use edTPA to make a summative judgment about educational consequences for the teacher candidate is reflected in the tempered response of just over 50% of the respondents reporting that they could make the same assessment of teacher candidate performance without an edTPA requirement. In sum, it appears that teacher educators are seeing the educative and learning value of the design of the activities within edTPA but are not as strongly committed to using the outcomes of the edTPA for consequential decisions about individual candidates, rather, tending to want to rely on their own judgements and other mechanisms for determining the quality of candidate performance.

Institutionally 63% of teacher educators experienced faculty resistance to implementing edTPA. Further, 40.5% reported that the implementation of edTPA had placed limits or interfered with the unique values and learning practices of existing curriculum at their institution. Since each program adopts edTPA in its own particular

way, this result raises questions about faculty participation and engagement with edTPA that, in turn, can affect the degree of support offered to teacher candidates. This is especially worrisome in that 31.5% of the respondent report that their institutions have not integrated edTPA within and across their licensure programs.

Table 4.2

*Survey results in educational consequences at program and institutional level.*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
the expectations and rubrics of edTPA are helpful in revising the learning content of coursework in our programs.	4.5% (1)	27.0% (6)	45.0% (10)	22.5% (5)
implementing edTPA helped our program to improve the learning experiences for our teacher candidates.	9.0% (2)	31.5% (7)	31.5% (7)	27.0% (6)
our teacher candidates' results on edTPA have helped us make decisions about the individual's readiness to teach.	22.5% (5)	36.0% (8)	31.5% (7)	9.0% (2)
implementing edTPA has taken away some of our autonomy to design our courses and programs.	18.0% (4)	40.5% (9)	27.0% (6)	13.5% (3)
I and my colleagues could make the same assessment of teacher candidate performance without edTPA requirement.	9.0% (2)	36.0% (8)	36.0% (8)	18.0% (4)
the implementation of edTPA has placed limits or interfered with the unique values and learning practices of existing curriculum at our institution.	9.0% (2)	49.5% (11)	40.5% (9)	0.0% (0)
implementing edTPA allows our faculty to see trends in our program impact on teacher candidates' performance.	4.5% (1)	22.5% (5)	58.5% (13)	13.5% (3)

all relevant faculty at our institution currently know about edTPA expectations.	18.0% (4)	40.5% (9)	22.5% (5)	18.0% (4)
our institution has experienced faculty resistance to implementing edTPA.	13.5% (3)	22.5% (5)	9.0% (2)	54.0% (10)
our institution has integrated edTPA within and across our licensure programs.	13.5% (3)	18.0% (4)	45.0% (10)	22.5% (5)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

**Meaningfulness.** This dimension investigated what aspect of edTPA created meaningful experiences to motivate participants' full participation. Teacher educators reported that contextualized tasks may create conditions where teacher candidates could be motivated and, at the same time, their teacher candidates also needed to feel some responsibility for taking this assessment seriously. In addition, teacher educators reported that engagement in edTPA led to meaningful experiences for faculty members by validating their current work and creating opportunities to collaborate with other faculty members.

**Classroom level.** Teacher educators reported that edTPA resembles real teaching practices and required tasks are authentic (81%). However, only 27% reported that their teacher candidates perceived this assessment as a worthwhile learning experience and only 36% reported that their candidates perceived edTPA as being relevant to the teaching context in which they completed it. These results seem contradictory in nature. If the assessment is viewed as authentic teaching, yet the experience is not a worthwhile experience or relevant to the context, then what would be viewed as a worthwhile learning experience or what teaching activity is relevant to the context? As a reminder, these results are how teacher educators perceive the views of their teacher candidates. We may actually be interpreting how teacher educators view edTPA in these reports, or we

may be seeing a result of how teacher educators presented edTPA to their candidates. That is, the edTPA tasks, in and of themselves are authentic to teaching and educative in value (as we saw in the discussion about educational consequences). Yet, as an assessment, the worthwhile nature and relevance of the tasks, is called into question. This is a difficult set of data from which to draw meaning without further in-depth discussion with teacher educators.

Table 4.3

*Survey results in meaningfulness at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
the task requirements of edTPA closely resemble real teaching practices.	4.5% (1)	13.5% (3)	54% (12)	27.0% (6)
our teacher candidates perceived edTPA as a worthwhile experience for preparing to be a teacher.	31.5% (7)	40.5% (9)	27.0% (6)	0.0% (0)
our teacher candidates perceived edTPA as being relevant to the teaching context in which they completed it.	18.0% (4)	45.0% (10)	36.0% (8)	0.0% (0)
our teacher candidates engaged with edTPA with rigor and good effort.	4.5% (1)	49.5% (11)	36.0% (8)	9.0% (2)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

Teacher educators viewed edTPA as a meaningful assessment for teacher candidates because it supported them in showing what they know and what they can do during classroom instruction. One teacher educator remembered a moment when she had conversations with a group of teacher candidates. They chatted about how much they deeply understood cultural relevant pedagogy and applied the theory while lesson planning and instruction as they began to write commentaries for edTPA. The required writing prompts supported teacher candidates to validate how they used their knowledge

in their instructional practice. The same teacher educator also observed that the teacher candidates acted differently and brought more stories from their clinical sites to faculty members and supervisors as they were engaged in edTPA.

However, from the teacher candidates' perspective, one teacher educator reported in an interview that "because there is not an enforced pass rate for edTPA in Minnesota, some candidates do not take it seriously." Another teacher educator shared her experience:

the biggest challenge for them or one of them is that it's not required; they don't have to pass to get their teaching license. And in the cohort model, where they're all talking to each other all the time, that message, although we don't say it – in fact, we're the counter to it, we're the ones saying, "Yes, but this is going to be similar to assessments that you'll have to do to get tenure as a teacher. When you want to defend like why you should be promoted it's going to look a lot like edTPA.

Since there was no immediate high-stakes consequence for teacher candidates, edTPA was not a high priority for all candidates when compared to student teaching and job searching. While the underlying theory of edTPA was recognized as authentic to teaching, the current status of it being a low-stakes assessment for teacher candidates discouraged teacher candidates from fully engaging in the learning process in edTPA. This interpretation was supported by survey data--54% of survey respondents disagreed that their teacher candidates engaged with edTPA with rigor and good effort.

**Program and institutional level.** As shown in Table 4, 76.5% of teacher educators reported that, as an institution, the results of edTPA are important for their

program improvement efforts. Additionally, 58.5% agreed that implementing edTPA contributes to their collaborative work across their institution.

Table 4.4

*Survey results in meaningfulness at program and institutional level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
implementing edTPA facilitated meaningful interactions between faculty and our teacher candidates.	13.5% (3)	27.0% (6)	45.0% (10)	13.5% (3)
implementing edTPA created more opportunities for faculty collaboration.	18.0% (4)	18.0% (4)	40.5% (9)	22.5% (5)
implementing edTPA in programs improved our overall program coherence.	18.0% (4)	31.5% (7)	22.5% (5)	27.0% (6)
our faculty perceived edTPA to be a worthwhile learning experience for our teacher candidates.	27.0% (6)	22.5% (5)	36.0% (8)	13.5% (3)
implementing edTPA contributes to our collaborative work across our institution.	18.0% (4)	22.5% (5)	36.0% (8)	22.5% (5)
as an institution, the results of edTPA are important for our program improvement efforts.	13.5% (3)	9.0% (2)	54% (12)	22.5% (5)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

In terms of the meaningfulness of edTPA for programs and institutions, teacher educators perceived that edTPA confirmed what they have worked to support teacher candidates' learning to teach in their programs and what they hope their candidates will learn while in their programs. One of them said:

...it certainly gave a lot of validity to the need for attention to academic language and to focusing on English learners. And so that's been a real benefit I see in terms of other program areas.

They agreed that their programs already had learning activities very similar to the learning opportunities for teacher candidates in edTPA such as writing reflections and doing student work analysis. However, they reported that this assessment gave the external validity to the kinds of learning opportunities for teacher candidates that they wanted to promote.

Additionally, 64% of teacher educators agreed that the implementation of edTPA created more opportunities for faculty collaboration, bringing faculty together around work that is meaningful to the program. An example of faculty collaboration was a partnership between two programs for learning academic language. One faculty member developed presentations to faculty members and teacher candidates for understanding academic language and how to use the concept in instructional planning. Faculty members reported that it was a great opportunity to learn with each other as it never happened before. In addition, it created more collaborations with new hired graduate supervisors. For example, one teacher educator reported:

So I'm going to propose a supervisor retreat in December, where, happy hour, whatever it takes, where the handbooks are on the table and we actually look at the rubrics and talk about it and come to some consensus about how they can then support the candidates in the field a little bit stronger. Because I think that would only be a good thing, just to make sure, just because we have some new supervisors this year, and they're coming from contexts where this is completely new for them. So that would be good.

Taken together, the ideas of validating existing practices and creating new opportunities for collaborating on curriculum and learning experiences are indicators of the



meaningfulness behind the adoption of the new performance assessment. These activities are not trivial in the everyday work of teacher educators. They reported that the assessment adoption triggered these new understandings about their practices and about how to work with their colleagues.

**Transparency.** This dimension investigated how all participants clearly understood the overall assessment system including rubrics, expectations, scoring procedures, support materials, and connections to learning subjects in existing curriculum. Teacher educators showed mixed perspectives about how to work with teacher candidates using this new assessment.

**Classroom level.** Faculty members were confident in their own understanding of this assessment system. Ironically, they perceived that their teacher candidates had less clear understanding of edTPA. For example, Table 5 shows that 76.5% of teacher educators surveyed reported that their teacher candidates had opportunities to practice the kinds of activities that are expected in edTPA before they submit their final assessment for scoring. Also 63% agreed that their institutions supported their teacher candidates' understanding of edTPA rubrics through course activities and assignments. However, only half of teacher educators agreed that their teacher candidates had a clear understanding of the rubrics and made a connection to their teaching in classroom. These results suggest that supports from faculty members and institutions may need to be re-examined in terms of how the current activities and supports effectively support teacher candidates in understanding the expectations of the assessment. We are reminded here that the candidate interviews suggested that they enjoyed the supports that were provided

to them, but wanted more support in understanding how to use the rubrics to analyze their teaching practice rather than feeling like they were guessing about the expectations.

Table 4.5

*Survey results in transparency at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
our teacher candidates have opportunities to practice the kinds of activities that are expected in edTPA before they submit their final assessment for scoring.	4.5% (1)	18.0% (4)	45.0% (10)	31.5% (7)
our institution supports our teacher candidates' understanding of edTPA rubrics through course activities and assignments.	13.5% (3)	22.5% (5)	31.5% (7)	31.5% (7)
our teacher candidates have a clear understanding of edTPA rubrics and how they relate to classroom teaching.	9.0% (2)	45.0% (10)	36.0% (8)	9.0% (2)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

The interviewed teacher educators presented mixed opinions about how much they should be involved in preparing teacher candidates for edTPA. One group of teacher educators actively supported teacher candidates with different strategies. One was to review handbooks with either instructor during coursework or supervisors during practicum seminar. This activity allowed teacher candidates to get familiar with the language and fundamental components of the assessment. In addition, the rubrics were used for developing coursework assignments and modifying lesson plan templates. An instructor reported that the idea of breaking teaching into planning, instruction, and assessment was a simple but precise way to describe a sequence of teaching.

Furthermore, the faculty members in two programs incorporated technology to support their candidates to get familiar with edTPA tasks. One program leader reported:

This year they used VideoAnt to do annotations on their videos. Then they write up the prompts for task two and submit to their supervisors for a grade. And that could be a video of any lesson...that's a way for them to get familiar with making a video, practicing gathering that information, as well as then reflecting on those prompts.

VideoAnt is a web-based annotation tool for mobile and desktop devices developed by the University of Minnesota. The teacher education faculty member reported that VideoAnt was a tool to support teacher candidates to practice real edTPA tasks of analyzing their teaching.

While interview data revealed that faculty members had tried different strategies every year to figure out the best supports for teacher candidates, one faculty member showed a strong standpoint that edTPA should be the sole responsible of teacher candidates. He viewed that faculty involvement in the assessment activity is a form of “teaching to the test.” He argued that the whole assessment process should be completed by teacher candidates individually without supports from teacher educators. This reaction showed faculty’s resistance toward edTPA. Similar reactions, were observed by the survey respondents in other institutions. In response to this issue, one faculty member said:

...if the faculty in my teacher education department were better versed in the components of edTPA (or saw a reason to be) their coaching of students in their clinical sites would take on a different tone and be more focused on planning, instruction and assessment.

This quote implied a significant role of faculty members to set a tone for how to introduce edTPA to candidates and help make the expectations more transparent through embedded coaching activities during the program.

**Program and institutional level.** Teacher educators reported that less than half of faculty have clear understanding of edTPA rubrics and other supporting materials including *Thinking Behind the Rubrics* and *Making Good Choices*. On the other hand, they viewed that teacher candidates and faculty members have a better understanding of the submission deadlines and official release dates. These survey results raise questions about how teacher candidates can gain transparent and deep understanding of the assessment expectations if their faculty's understanding of the assessment rubrics and other accessible documents is not clear.

Institutionally, teacher educators viewed that institutions are predominantly embedding edTPA into curriculum, systematically introducing candidates to the requirements, and ensuring that candidates understand the scoring process. This result implies that the mechanism of the assessment implementation seems to be taking root institutionally.

Table 4.6

*Survey results in transparency at program and institutional level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
our faculty have a clear understanding of edTPA rubrics.	9.0% (2)	45.0% (10)	40.5% (9)	4.5% (1)
our faculty have a clear understanding of other supporting materials (e.g., <i>Thinking behind the rubrics</i> , <i>Making good choices</i> )	13.5% (3)	45.0% (10)	40.5% (9)	0.0% (0)

our teacher candidates and faculties fully understand the timeline of submission deadlines and official score release dates.	9.0% (2)	18.0% (4)	54.0% (12)	18.0% (4)
our institution has intentionally embedded edTPA or related activities within our programs.	9.0% (2)	22.5% (5)	31.5% (7)	36.0% (8)
our institution systematically introduces all our teacher candidates to edTPA and its requirements.	4.5% (1)	31.5% (7)	27.0% (6)	36.0% (8)
our institution ensures that teacher candidates understand how edTPA will be scored.	4.5% (1)	22.5% (5)	40.5% (9)	31.5% (7)
our institution has a local policy for how we use our teacher candidates' performance on edTPA based on local scoring and/or formal score reports.	9.0% (2)	31.5% (7)	31.5% (7)	27.0% (6)
our teacher candidates have a clear understanding of how edTPA formal results will or will not be used by our institution	9.0% (2)	27.0% (6)	45.0% (10)	18.0% (4)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

Two interviewed faculty members raised a concern that it was necessary to establish supporting systems for teacher educators who were not familiar with edTPA or were newly hired.

**Directness.** This dimension investigated how results can be immediately interpreted by participants with little need for technical translation or indirect interpretation. A minimum amount of interpretation should be place in order to be able to use results for learning and teaching. By observing teacher candidates' performance and reviewing official scores from external scorers, teacher educators were able to make decisions for program improvement and institutional policy. Yet, teacher educators did not have a strong understanding about how feedback related to edTPA directly scaffolded teacher candidate's learning how to teach.

**Classroom level.** Teacher educators confirmed that participating in edTPA has the potential to generate educative feedback for teacher candidates. Table 7 shows that 40.5% of teacher educators surveyed reported that teacher candidates used feedback from instructors related to edTPA performance to improve their teaching. The edTPA design and supporting materials were viewed as sources of evidence for providing local feedback by 82% of the respondents, suggesting that a close and direct link between the assessment and candidate development exists in the eyes of these teacher educators. One teacher educator suggested that the kind of feedback that teacher candidates get while they are completing edTPA is better than how current teacher evaluation systems are set up, suggesting strong connections between edTPA and learning how to teach:

...it's not the exact same teacher evaluation system that the schools or districts are using. In some respects, it's better because it's content specific, where many schools and districts are using teacher evaluation tools that are more general.

However, in terms of formal score reporting and its direct use in improving performance, only 22.5% of the respondents thought that teacher candidates knew how to use the formal score reports of edTPA to plan early career development and similarly, and only 22.5% agreed or strongly agreed that candidates knew how to use their edTPA scores to plan for their induction into the profession. So, while the assessment and supporting materials have the potential to help make direct connections between the assessment and coaching and the development of teaching performance, the formal score reports are not yet viewed as helpful in these formative phases of teacher development.

Table 4.7

*Survey results in directness at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
edTPA design, expectation and other supporting materials provide enough guidance that I could give specific local feedback to our teacher candidates on their teaching performance during their teaching preparation program.	9.0% (2)	9.0% (2)	67.5% (15)	13.5% (3)
our teacher candidates use feedback from instructors related to their edTPA performance prior to official scores to improve their teaching.	13.5% (3)	45% (10)	40.5% (9)	0% (0)
our teacher candidates understood their edTPA formal score reports.	22.5% (5)	22.5% (5)	54.0% (12)	0% (0)
our teacher candidates know how to use their formal score reports on edTPA to plan their early career development as a teacher.	31.5% (7)	45.0% (10)	22.5% (5)	0% (0)
our teacher candidates know how to use their formal score reports on edTPA to plan for induction.	31.5% (7)	45.0% (10)	18.0% (4)	4.5% (1)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

One faculty member reflected that reviewing edTPA results supported a quick program evaluation by using individual teacher candidate' scores. He began his review by looking at each rubric. Some rubrics had more level ones (the lowest score) and other rubrics had more level fives (the highest score). And then he found teacher candidates who got a one or a five on particular rubrics. This process helped him to make sense of how his program supported its teacher candidates and how much they had learned throughout the program. He viewed that this quick reviewing process helped to look at a big picture of his program and how to plan for the next cohort.

One of the issues that surfaced during the teacher educator interviews related to the direct use of edTPA for improving performance related to the timing of the assessment administration. An administrator of edTPA at one institution reported feeling “a real-time crunch” to give feedback to teacher candidates so that they could improve their teaching. They experienced a schedule conflict between the student teaching site and university schedule. Typically, teacher candidates begin to record a video and write reflections during the last month of student teaching. It is a time when teacher candidates emerge as the full-time teacher, taking over a whole class from the cooperating teacher. Program deadlines for submitting the edTPA were set up in the same month. This timeline did not support teacher candidates in using the formal assessment process and resulting scores formatively.

Among the institutions in this study, official scores are typically returned to teacher candidates after they graduate from the programs. So, it is really unknown how teacher candidates use edTPA results for the first year of teaching. Similar issues exist across institutions in term of timeframe, one faculty member proposed an alternative approach to embedding this assessment within programs:

I believe if we required our student teachers to complete their edTPA sooner, during their clinical semester, we'd have the scores back before the end of the semester and be able to use them to lead self-reflection, problem solving, and professional growth plans for their induction year in teaching. We have so much to do in these areas.

This statement raised a question of when the best time for administering edTPA would be in order to best support teacher candidate' learning.



**Program and institutional level.** For programs and institutions, directness of the assessment is bound up in how the programs can use the data from the assessment for program improvement. The survey questions investigated how well faculty could read, interpret, and use results of the assessment for decision-making and improvement. The survey responses show that there was a fair amount of agreement that the respondents and their colleagues could read and interpret the assessment results, that the rubrics show areas of program strength and areas for improvement, and that the institution uses the results to make decisions about program improvement.

One of the survey items with the least agreement showed only 58.5% of respondents agreeing or strongly agreeing that edTPA results are based on evidence that I think is a good representation of their teaching. This result would suggest that much program improvement is taking place based on data that the respondents think is not a good representation of candidates' teaching. These program faculty are finding use of the data, but if the assessment results are not viewed as strong representation of candidates' teaching, some inferences are happening during the interpretation process. Therefore, it is not clear that respondents are finding a direct use of the data for program improvement, but rather an indirect use of the data for program improvement.

Table 4.8

*Survey results in directness at program and institutional level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
I know how to read and interpret edTPA results for our programs.	4.5% (1)	13.5% (3)	54.0% (12)	27.0% (6)
our program employs local evaluations to provide formative feedback to our teacher candidates.	18.0% (4)	27% (6)	31.5% (7)	22.5% (5)

edTPA results on individual rubrics show me where our program(s) have strengths and where there are areas for improvement.	4.5% (1)	18.0% (4)	45.0% (10)	31.5% (7)
edTPA results for our teacher candidates are based on evidence that I think is a good representation of their teaching.	18.0% (4)	22.5% (5)	36.0% (8)	22.5% (5)
our institution has a process for sharing and examining edTPA scores for licensure programs.	13.5% (3)	9.0% (2)	54% (12)	22.5% (5)
overall, our faculty know how to read and interpret edTPA results.	18.0% (4)	13.5% (3)	58.5% (13)	9.0% (2)
our institution uses the results of edTPA to make decisions about program improvement.	13.5% (3)	13.5% (3)	58.5% (13)	13.5% (3)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

In interviews, examples of how edTPA data provided evidence to make decisions for program improvement were provided. One program leader described that her institution reviewed edTPA official scores across program areas the previous summer. There was a pattern that scores in assessment domains were much lower than planning and instruction domains. The scores gave a general indication of what should be examined in the program curriculum, but only indirectly. The faculty reviewed selected teacher candidates' portfolios to get a more direct sense of candidates' performance to inform the current coursework across programs. This review revealed that the teacher candidates failed to make a connection between learning objectives and the assessments they implemented. Also, they didn't clearly state an intention of the student assessment in reflecting how their feedback would improve students' learning. These observations of the candidates' actual edTPA tasks, not only the aggregate program scores, lead to a change of curriculum across the institution, giving special attention to classroom

assessments. The participant explained that this process created an opportunity to share concerns and to learn about different educational strategies in different programs.

**Fairness.** This dimension investigated how teacher candidates were given sufficient opportunities to demonstrate their teaching ability and potential. The survey and interview questions focused on whether candidates had equal opportunities to learn about the assessment and whether candidates were held to the expectations of the assessment only and not something that was not part of the assessment.

**Classroom level.** There was consensus that teacher educators fully understood their ethical positions throughout the assessment activities with 90% of teacher educators reporting that they felt that their faculty follow the ethical coaching expectations for edTPA. Also, 81% of the respondents viewed that their teacher candidates received appropriate mentoring and coaching for preparing their edTPA. There was a significant fairness concern suggested in these survey results with regard to writing: 85.5% indicated that the writing demands of the assessment were challenging for some of the students. Interview data mirrored these challenges in their classroom as well.

Table 4.9

*Survey results in fairness at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
our faculty follow the ethical coaching expectations for the edTPA.	4.5% (1)	4.5% (1)	45.0% (10)	45.0% (10)
our teacher candidates received coaching or mentoring for preparing their edTPA.	9.0% (2)	9.0% (2)	54.0% (12)	27.0% (6)
when we locally support edTPA, we stay focused on the expectations for teaching described in the rubrics.	4.5% (1)	27.0% (6)	49.5% (11)	18.0% (4)

the writing demands of edTPA are challenging for some of our teacher candidates.	0.0% (0)	13.5% (3)	36.0% (8)	49.5% (11)
--	-------------	--------------	--------------	---------------

---

*Note: Due to rounding, percentages do not appear to add up to 100%.*

Interview data showed that teacher educators attentively followed ethical guidelines to avoid causing issues around fairness. The educators read ethical coaching guidelines multiple times to confirm their understanding of their ethical responsibilities and to identify potential sources which may raise issues. At the same time, teacher educators experienced difficulties applying the guideline when they were mentoring and coaching teacher candidates. They had a dilemma in working with a group of students who struggled more than others. Even though struggling teacher candidates sought out more direct and specific feedbacks and comments, teacher educators tried not to give too specific feedback to meet ethical coaching guidelines. One teacher educator felt that a requirement of ethical coaching constrained coaching and support for teacher candidates based on individual's needs.

All interviewed teacher educators except two, raised fairness issues related to extensive writing requirements of edTPA. They observed that teacher candidates struggled to develop and organize their written commentaries. One faculty member mentioned:

Part of my concern is that candidates have knowledge and skills but they are weak in their ability to reflect and therefore responses to commentary prompts either don't directly answer the question or don't provide enough information.

It was a little surprising to this teacher educator that teacher candidates didn't have strong professional and academic writing skills. The faculty member didn't generalize this struggle to all teacher candidates, but he assumed that teacher candidates who had more

experience in writing would be less likely to struggle with the assessment because of the writing expectations.

...they find the most difficult part is writing and feeling like they've already answered the question once, and then they write some more and they feel like, "I'm going in circles." And that's just because they're probably not as accomplished writers as maybe the English ed people are.

Similar assumptions were common among the participating teacher educators. A faculty member from another program said, "it is an assessment that is very writing heavy, so if you're intimidated by writing, you might go into it with more anxiety." She pointed out that that even if the writing does not directly count toward the score, the emotion may discourage teacher candidates to perform accurately. Each program had different perspectives on how to handle this issue, yet some programs planned to send teacher candidates who needed more supports in writing elements to the campus writing center in the future.

**Program level.** Supervisors and cooperating teachers play significant roles for teacher candidates when they are learning to teach. For the candidates to have a fair experience with edTPA, we would expect that the educators surrounding the candidates have been adequately and equally prepared to support the candidates. In the survey, 72% of teacher educators reported that their programs worked with supervisors to help them understand how to support teacher candidates while they are student teaching while only 49.5% of them worked with cooperating teachers to prepare for supporting teacher candidates. Given such statistics, there was a gap in the preparation of cooperating teachers that could influence the fairness for supporting teacher candidates on edTPA.

Table 4.10

*Survey results in fairness at program level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
we expect our teacher candidates to put their best effort forward on their edTPA submission.	0.0% (0)	4.5% (1)	45.0% (10)	49.5% (11)
our program works with clinical supervisors to help them understand how to support the teacher candidates with edTPA while they are student teaching.	9.0% (2)	18.0% (4)	49.5% (11)	22.5% (5)
our program works with cooperating teachers to help them understand how to support the teacher candidates with edTPA while they are student teaching.	9.0% (2)	40.5% (9)	49.5% (11)	0.0% (0)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

University faculty understood that an involvement of cooperating teachers in edTPA implementation would be beneficial to teacher candidates. Yet, there was a hesitation to ask cooperating teachers to take active roles for supporting teacher candidates' edTPA at the program level. A faculty member said:

I've always taken sort of a little bit of an approach of I don't want the cooperating teachers to get mired by this. And so we send out a letter to the teachers explaining what it is, and that this is going to happen, but it really is your teacher candidate's responsibility. I'm afraid that if we try to bring the cooperating teachers in from a clinical instructional position, that would just take so much training to get it right. I'd rather they just sort of keep their nose out of it.

Other faculty members from different programs shared a similar perspective about asking cooperating teachers to be part of edTPA implementation, although some cooperating teachers volunteered to take active roles to support edTPA. However, faculty members

who coordinate programs did not want to put too much control in the hands of the student teaching site and cooperating teachers. This perspective has the potential of creating less-than-fair circumstances for teacher candidate' in terms of how some might be more supported if cooperating teachers are trained on edTPA expectations and others are not. As indicated above, the candidates who had cooperating teachers who helped them to be planful about edTPA evidence collection while student teaching found this to be a beneficial part of their overall edTPA support.

**Institutional level.** Across institutions in Minnesota, 67.5% agreed or strongly agreed that their institutions ensured that teacher candidates had an equitable opportunity to prepare their edTPA. However, 81% of teacher educators surveyed viewed that their institution provided appropriate technology supports for teacher candidates to complete edTPA.

Table 4.11

*Survey results in fairness at institutional level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
our institution ensures that teacher candidates have an equitable opportunity to prepare edTPA.	13.5% (3)	18.0% (4)	58.5% (13)	9.0% (2)
some teacher candidates struggle with the process of uploading and submitting edTPA in an electronic platform.	0.0% (0)	18.0% (4)	63.0% (14)	18.0% (4)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

In consideration of fairness at the institutional level, teacher educators were concerned that edTPA itself may cause inequality as each institution has different capacity to prepare teacher candidates for edTPA implementation. One faculty member stated the following:

...our students are 50% students of color and have difficulty completing some of the tasks. The edTPA is another example of how institutional racism and the need for tests, regulation, hurdles, and roadblocks to be established so individuals from unique backgrounds cannot succeed and or move up our system.

This argument was mirrored by public concern in which particular groups of students are underperforming on edTPA because of personal situations including SES, racial and ethnic background. Another faculty member explained that a constraint of financial and personnel resources caused this issue since there were not sufficient funds and personnel to support each teacher candidate based on their needs. On the other hand, another faculty member suggested a constructive and reimagining approach. She said, "...open up the practice of what professors do in their university classrooms and there is less behind 'academic freedom.'" This quote implies that a role for teacher educators is to be responsive to creating fair assessment sites for their students.

**Usability.** This dimension investigated how this performance-based assessment was experienced as a practical and sustainable tool in the situated context of teacher education programs.

**Classroom level.** By observing cases of edTPA implementation in other states, one of the major concerns was time demands on faculty members for preparing their teacher candidates for edTPA and for integrating the assessment into existing curriculum. From the survey data, only 22.5% of teacher educators responded that the time teacher candidates spent on edTPA was reasonable in light of the other tasks and responsibilities. However, 72% of the same group of teacher educators viewed that this assessment is useful for determining their teacher candidates' readiness to teach. So, there is the



possible interpretation that the time spent is worth it to the teacher educators completing this survey. In addition, it was hard to avoid additional instructional time for preparing teacher candidates for the logistical processes of the assessment. Table 12 shows that 81% of teacher educators reported that time was dedicated to teach electronic submission platform, teaching writing skills, and video editing. Only 58.5% viewed that these logistical processes were useful skills for teacher candidates to learn.

Table 4.12

*Survey results in usability at classroom level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
the amount of time our teacher candidates spend on edTPA is reasonable in light of the other tasks and responsibilities they have.	22.5% (5)	54% (12)	22.5% (5)	0.0% (0)
even with the additional workload of edTPA, it is a useful assessment for determining our teacher candidates' readiness to teach.	13.5% (3)	13.5% (3)	49.5% (11)	22.5% (5)
implementing edTPA has required adding or using instructional time for logistical processes such as learning an electronic submission platform, teaching writing skills, video editing.	9.0% (2)	9.0% (2)	40.5% (9)	40.5% (9)
the logistical skills needed for edTPA (e.g., electronic platform for submitting edTPA, writing, video recording classroom) are useful skills for our teacher candidates to learn.	13.5% (3)	27% (6)	58.5% (13)	0.0% (0)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

Overall, the survey results confirmed that implementation of edTPA demanded instructional time both for teacher candidates and teacher educators. Interviewed teacher educators echoed this concern. However, teacher educators who have experienced with

edTPA closely were less concerned about time demands. One faculty member, who was involved in edTPA implementation from the first year, remembered that teacher educators intuitively expressed concern about time, effort, and energy demands for working on the assessment. However, she observed that the stress level created by the demand of time decreased after pilot experiences over three years. Her observation was supported by another interviewee who suggested that edTPA hadn't fully been embedded in her program and it generated more time demands for teacher candidates. The faculty member said, "we [faculty members in her program] didn't want to give up the pieces of our program we already felt were strong." As she and her colleagues decided to add edTPA into their existing curriculum, it turned out to be extra work for teacher candidates. However, she experienced a perspective shift on how to use edTPA as she had gained more familiarity with edTPA. Still, she believed that there is a gap between their existing curriculum and edTPA components. Yet, she was planning to have a faculty meeting over the summer to adopt more edTPA components into their curriculum. She expected that if they can find a way to embed edTPA without losing the strengths in the program curriculum, it would reduce stress of the teacher candidates. Based on this instance, it is a fair interpretation that the initial investigation of time, efforts, and energy is hard to avoid. Yet, the stress and over-demands on time in the curriculum does not need to persist continuously.

**Program and institutional level.** At the program level, there were mixed perceptions about efficiency and accuracy of this assessment. Just half of the teacher educators agreed that this assessment provided an efficient process for assessing teacher candidates' performance in their licensure programs. Additionally, Table 13 shows that

54% of them responded that edTPA provided an accurate measure for teacher candidates' performance.

In terms of paying edTPA fees, 82% reported that their institutional process for paying the fees were equitable for all teacher candidates and 45% viewed that paying edTPA fees created unreasonable financial burden for their teacher candidates. The financial burden is widely recognized as having negative side-effects for edTPA implementation.

Table 4.13

*Survey results in usability at program and institutional level*

Survey Questions	Strongly Disagree	Disagree	Agree	Strongly Agree
edTPA provides an efficient process for assessing teacher candidates' performance in their licensure program.	27.0% (6)	22.5% (5)	31.5% (7)	18.0% (4)
edTPA provides an accurate measure of the teaching performance of our candidates.	13.5% (3)	31.5% (7)	49.5% (11)	4.5% (1)
the amount of time our faculty spends on tasks related to the edTPA is reasonable in light of the other tasks and responsibilities they have.	22.5% (5)	40.5% (9)	31.5% (7)	4.5% (1)
the investment our institution has made in preparing faculty for the implementation of edTPA is a good investment in our teacher education programs.	18.0% (4)	22.5% (5)	40.5% (9)	18.0% (4)
our institution's process for paying edTPA fees is equitable for all candidates.	18.0% (4)	9.0% (2)	45.0% (10)	27.0% (6)
our institution's process for paying edTPA fees creates an unreasonable financial burden for teacher candidates.	13.5% (3)	40.5% (9)	22.5% (5)	22.5% (5)

*Note: Due to rounding, percentages do not appear to add up to 100%.*

Interviewed teacher educators pointed out a few issues in using this assessment and proposed suggestions for making the assessment more practical. One of them was to complete the edTPA requirement at the beginning of student teaching, so that teacher educators would have a better understanding of their teacher candidates at the starting point of student teaching. One teacher educator suggested:

if we required our student teachers to complete their edTPA sooner during their clinical semester, we'd have the scores back before the end of the semester and be able to use them to lead self-reflection, problem solving, and professional growth plans for their induction year in teaching.

With the current implementation, edTPA official scores are returned to the institution after teacher candidates leave their programs. This time structure allowed teacher candidates to develop a sense that edTPA was not to be taken seriously as they left their program before scores were not yet available to the program faculty. While some programs counted edTPA official scores in the final grade of the seminar class during student teaching, it did not create a huge difference to teacher candidates by combining with other scores in that class.

Another suggestion was to focus on selected areas for each teacher candidate. One clinical supervisor thought that the expectations of teaching in edTPA are overwhelming to pre-service teachers. This person observed that a feeling of being overwhelmed created a negative attitude toward this assessment among teacher candidates. This teacher educator proposed that if teacher educators or teacher candidates were allowed to select areas to focus and plan for learning, it would be a more practical and formative tool for

teacher candidates rather than pressing teacher candidates to a feeling of being overwhelmed.

Lastly, financial burdens on candidates were raised during interviews. This concern was widely recognized as having negative side-effects for edTPA implementation. Yet, teacher educators haven't provided specific examples for how much teacher candidates struggle to pay fees.

### **Summary of Findings**

Based on findings from interviews and surveys of teacher educators and teacher candidates, this chapter answered the research question: How can a consequential validity framework help describe the implementation of a standardized complex performance assessment administered in local contexts?

Six dimensions of a consequential validity framework contributed to capture what happened during an assessment implementation in a field. Overall, teacher candidates and teacher educators viewed that edTPA created opportunities to examine and improve their current practice in teaching and learning. These are evidence of positive consequences which aligned with anticipated outcomes. For teacher candidates, a reflective practice was a key area of learning as they participated in this assessment. For teacher educators, edTPA was an opportunity to map out current curriculum and to evaluate it for program improvement. At the same time, both groups identified negative consequences which were not anticipated by test developers. Evidence of negative consequences may be critical information for better assessment implementation to support the educative purpose of this assessment.

As Minnesota decided not to use edTPA results for licensing approvals, the test results were more likely to be used for program improvement. For programs, edTPA test results served as a validation of existing curriculum. One of the major institutional decisions during the adoption of edTPA in Minnesota was mapping out current curriculum to find the alignment with edTPA. This process required faculty to review and develop a better understanding of their curriculum. Another decision was to integrate academic language into instruction and curriculum. The concept of academic language has been studied and discussed in selective discipline areas, yet not all faculty were familiar with the concept and how to integrate it in teacher preparation programs. As this concept was introduced to faculty by edTPA, it encouraged them to think deeply about integrating academic language into content-specific teaching.

Both teacher candidates and teacher educators recognized a theoretical significance of edTPA to support teacher candidates' development of their teaching. However, the two groups of interviewees identified significant contextual and circumstantial influences on their efforts to convert theory to practice in their classrooms, programs, and institutions. edTPA was designed with a promise that an authentic assessment tool would support learning by practice and to enhance motivation to participate in assessment. Yet, it turned out that meaningfulness of this assessment was established by participants. It depended on how this assessment was taken up by faculty members and embedded in existing curriculum. This dimension implies significant roles of participants to create positive local contexts for the assessment to take on meaningfulness for both the test takers and the test administrators.

In terms of transparency, this performance-based assessment has been administered without deep understanding held by every teacher educator and teacher candidate. While the technical and logistical aspects of the assessment seemed to be readily understood, the deeper meaning of the expectations for teaching as described in the rubrics and other supporting materials were not yet part of everyone's understanding. Under this circumstance, it is a fair interpretation that both teacher educators and teacher candidates had difficulty with the assessment experience and some candidates struggled to achieve anticipated goals of the assessment. Also, a lack of transparency caused misguided assumptions about the expectations of the assessment itself.

Teacher candidates liked to receive actionable feedback from teacher educators as a part of assessment activities. They viewed feedback on their teaching as direct evidence for understanding their current practice and what they needed to do to change for improvement. Teacher educators were divided in that some chose to use the assessment activities as learning experiences that led to the final assessment while others took the stance that the assessment was their candidates' sole responsibility and chose to not provide direct feedback on practice activities aligned with the assessment. Since reports of official score were scheduled after teacher candidates' graduation, teacher educators did not yet have a practice for how to use official score for feedback to their candidates. In terms of using the assessment as direct feedback for program improvement, teacher educators tended to use aggregated test scores to examine patterns of a group of teacher candidates, with one example of a teacher educators digging deeply into the candidate portfolios as a means to gathering evidence for program improvement.

A dimension of fairness was dominantly discussed in terms of ethical responsibilities to provide unbiased supports to all teacher candidates from the teacher educator's viewpoint. On the other hand, teacher candidates desired to have faculty's supports to prepare individualized preparation for edTPA. Teacher candidates argued for the critical role of teacher educators, including cooperating teachers and supervisors, to guarantee support for their performance while they navigated multiple local contextual factors. This dimension revealed a perception gap in responsibility of teacher educators between teacher candidates and faculty members.

The usability dimension revealed that time was a major concern for both teacher candidates and teacher educators to participate in this assessment. Both teacher candidates and teacher educators were challenged to create time for understanding this assessment in a light of other roles and responsibilities. Financial burdens are briefly discussed in literature, this study didn't find strong evidence related to financial burdens.

In sum, the framework captures implementation issues ranging from impacts on learning, to practice aspects of administration, to usefulness of the data and assessment results. The framework contributes to our ability to provide a comprehensive description of the implementation process



## CHAPTER 5

### DISCUSSION AND FUTURE RESEARCH

#### **Toward a Framework of Consequential Validity for Performance-Based Assessment**

Using the case of edTPA in Minnesota, the present study explored a concept of consequential validity for performance-based assessments. More often, validity studies have considered intended and unintended consequences as critical evidence to evaluate the consequential aspects of assessment. Breaking from this traditional practice, the framework used in this study examined assessment processes and activities during the adoption and implementation of the assessment in order to explore the local impact not only on the students, but also the consequential impact on the instructors, the curriculum, and the institutions where the assessment was adopted. This investigation provides in-depth explanations for how and why intended and unintended consequences are generated within local contexts.

The theoretical framework in this study reflects the dynamic and real context of assessment sites to account for how the nature of the assessment site supports and impedes the integration of the assessment and the use of assessment results. Grounded in findings in each dimension reported in chapter four, this chapter summarizes each dimension within the framework to further develop our understanding of the concepts underlying each dimension. Also, this chapter provides implications for future implementation of performance-based assessments. Additionally, a few limitations are discussed to address further research directions on performance-based assessments and their validity. Lastly, conclusions are drawn on key implications for further research studies.

As a reminder to the reader, this study explores the following two questions.

1. What constructs should be included in a framework that describes the consequential validity of an assessment?
2. How can a consequential validity framework help describe the implementation of a standardized complex performance assessment administered in local contexts?

In the next section, I answer the first research question (the second question was explored previously in chapter four), delving into each construct that I proposed at the beginning of this study and evaluate its usefulness within a framework for understanding consequential validity of performance assessments.

### **Six Dimensions of Consequential validity**

#### **Educational Consequences**

Many performance-based assessments aim to not only measure students' performance, but also provide learning opportunities for students and teachers. Desirable consequences are that performance-based assessments lead to positive effects by providing an opportunity for examining current practices and adjusting practices to achieve the goals set out in the assessments. So, this dimension investigates whether and how the performance-based assessment leads to positive and negative consequences for participants.

In Chapter Four, we saw that the edTPA, as a performance-based assessment, supported each group of participants differently. For teacher candidates, reflections on teaching activities supported them to revisit their own teaching practices and to think intentionally about how to improve their practices. By using a collection of teaching evidence, the assessment asked teacher candidates to reflect on their teaching—what was

going well and what needed to be improved in the dynamic contexts of teaching. For teacher educators, edTPA supported them to review the current support system and environment in which the performance assessment took place and consider how to improve educational supports that align with the assessment. Different from standardized assessments, students and their teachers' decisions create unique contexts for the assessment. The establishment of the assessment system involves learning the expectations and practicing the assessment tasks. To do so, it is necessary to carefully examine the current curriculum and to identify available resources, reasonable contexts, and the appropriate time in the curriculum sequence for supportive activities. Without fully understanding students' needs and contexts, it is difficult to expect positive and intended consequences of performance-based assessments.

The current perspective of validity in the testing and measurement community is fundamentally grounded in the assessment development phase. The traditional perspective does not pay attention to practices of implementing the assessment. Given that performance assessments such as edTPA are embedded in ongoing practices of teaching and learning, it is difficult to conclude that factors unanticipated by the assessment designers are irrelevant for evaluating the validity of the assessment. I believe that the traditional narrow scope of validity constrains our understanding of the implementation and administration of performance-based assessment in a field. Different from traditional assessments, context-based performance assessments require more flexibility with consideration of natural characteristics and contexts of performance-based assessment administration. Educational consequences in this study are clearly part of the assessment adoption and implementation process in local contexts, both in terms of the

impact on the student learning outcomes and in how the educational contexts shifted and changed in order to accommodate the new assessment.

Other dimensions including meaningfulness, transparency, directness, fairness, and usability in this framework help identify further evidence of positive and negative consequences on the assessment takers and the implementing institutions.

### **Meaningfulness**

Some psychometricians may claim that local implementation of performance-based assessments may be too loosely standardized, thus decreasing validity of assessment. However, the real-world contexts are not controlled in performance assessments. Rather, we view the dynamic and unique contexts as aspects of performance assessment that supports construct validity, assuming that the assessment is closer in proximity to the authentic practice that is being assessed. Authentic assessment increases opportunities to genuine engagement of students and educators so that generated data represent accurate information of students' achievement.

A dimension of meaningfulness investigates participants' perception of whether it is worthy to participate in an assessment. Participants engage in an assessment actively and genuinely when assessment activities are connected to their personal interests and values. Contextualized assessments are expected to create meaningful experiences as participants have a space to construct assessment responses that relate to their local work and interests. Furthermore, investigations of students' and teachers' understanding of performance-based assessments can offer important and relevant information necessary for understanding their perceptions and response to assessments. While the literature emphasizes meaningfulness in terms of the assessment takers' perception of the

assessment design, it is also important to recognize the role of the assessment settings in creating and supporting perceptions of the meaningfulness of the assessment. For example, assessment administrators and educators play a role in the creation of initial perceptions of learners and meaningful experiences.

The findings in chapter four revealed that teacher candidates showed different levels of attention to the performance-based assessment, depending on how they had been introduced to and taught about the tasks within the assessment. When provided hands-on practice through whole or partial assessment tasks, teacher candidates were more likely to engage and become aware of the meaningfulness of the assessment with regard to the actual practice of teaching.

Teacher candidates recognized the value of the assessment tasks at different points in time from various aspects of the assessment and from interactions with others. Some teacher candidates understood and recognized meaningfulness of the assessment tasks during practice with those tasks. Others found value in the tasks after they completed all requirements and paused to reflect on the overall experience as a learning process. Positive and negative attitudes to the assessment demonstrated by peers and educators also influenced assessment takers' engagement and self-reported effect on learning. The awareness of values encouraged the candidates to engage in the assessment actively and genuinely. Teacher candidates who viewed assessment activities as a valuable learning process were also less likely to prioritize the efficiency and financial burden of the assessment within the usability construct. When teacher candidates gave efficiency less priority, they seemed to have more positive educational experiences during the assessment.

The findings confirmed a strong association between meaningfulness and the transparency of the assessment purpose and uses (discussed in the section on transparency later in this chapter). As participants had a clearer understanding of the assessment and better insights of intended uses and purpose, they were more likely to find meaningful experiences with the assessment.

For large-scale assessments, low performance due to a lack of motivation is problematic (Penk, Pöhlmann, Roppelt, 2014). Even though contextualized tasks were expected to motivate teacher candidates' learning, it was important to make a personal connection with the assessment and the participants. It is crucial that the participants understand the details and expectations of a performance-based assessment.

Administrators and educators should pay attention to their attitude and perceptions of an assessment, which can have significant effects on how the assessment takers come to understand the meaningfulness of the performance assessment activities.

### **Transparency**

For successful implementation of performance-based assessment, all participants should have a clear understanding of the overall assessment process. While the reviewed literature has emphasized the importance of providing details of assessment procedures, the findings in this study suggest that participants need to be transparent about how this assessment is related to learning processes that are supported by the assessment. When performance-based assessments are implemented in an educational setting, the dimension of transparency is not limited to an understanding of the details of assessment administration, as the participants should also have a clear understanding of how the assessment is integrated into the local learning contexts. This level of transparency helps

alleviate and overcome local concerns regarding the disruption of normal routines and expectations.

Often performance-based assessments are standardized and structured with selective administrative instruction, tasks, rubrics, and scoring process. However, this aspect of performance-based assessment does not mean that all participants will use the assessment in the same way. Participants should be able to interpret the contexts carefully to implement performance-based assessment with a consideration of pre-existing curriculum. All users—instructors and students—should clearly understand that performance-based assessment is not just an add-on at the end of a program or instruction; instead, it should be embedded or integrated into the existing curriculum and instruction in order to support the development of the knowledge and skills that will be assessed.

This dimension is clearly the most critical dimension to support other dimensions. A lack of understanding of the assessment limits the participants' ability to identify potential factors associated with intended and unintended consequences. A lack of understanding in assessment design and process demands additional time to study the assessment itself. As test-takers are expected to complete all tasks within limited time, they rush to complete the assessment without fully engaging the assessment activities. Under this situation, it is difficult for test-takers to recognize learning opportunities as a value of the assessment. Additionally, it prevents administrators from navigating the contextual and personal factors to create positive and supportive assessments for test-takers.

## **Directness**

A dimension of directness examines how a performance-based assessment offers evidence of a process of learning and produces actionable feedback to learners. As performance-based assessments are situated in real-world contexts, we expect to give more individualized and actionable feedback to improve learning by using learners' own working samples as well as the external assessment scores.

The findings in this study examined how teacher candidates and teacher educators have used edTPA to support learning and teaching within teacher education programs. Limited evidence was available to investigate the kinds of feedback given to the teacher candidates while engaged in the edTPA assessment. However, teacher candidates agreed that this assessment had potential to provide direct evidence to show their readiness to teach and to support them in improving teaching by using samples of work from their teaching practice. Yet, they doubted how much they could ask questions and get feedback from faculty members during the final edTPA.

From teacher educators' perspective, official scores were direct evidence to review their students' achievement along with other data. This data review process supported educators and programs to recognize patterns of students' strengths and weakness. Learnings from this review were significant information to revise and improve curriculum and teaching for future students, but not as trusted for making decisions about the readiness to teach for individual teacher candidates.

This dimension is closely related to the dimension of educational consequences. In terms of connections with educational consequences, feedback is expected to create



learning opportunities for students. This immediate and individualized feedback from instructors and coaches is expected to create impactful learning moments.

Often numeric scores are treated as the only information from an assessment. It is an efficient and easy tool for decision making. However, numeric scores contain limited information about how and why a student obtained high or low scores. Assessments based on performance show thinking process for how they chose artifacts and constructed responses. It contains more information to make a link to instruction for learning improvement. Compared to other dimensions, this dimension gains less attention from both sides of test developers and users. However, if we claim an educative purpose of assessments, this dimension should be included in the framework of consequential validity as a qualitative distinction for how well the assessment evidence provides information for ongoing improvement.

### **Fairness**

Fairness aims to evaluate how a performance-based assessment facilitates equitable opportunities for learners to demonstrate their potential through performance. This dimension expands the scope of fairness within the tradition perspective of test design. A traditional perspective assumes that the same tasks sampled from a range of relevant topics and development of reliable and systematic scoring procedures guarantee fairness of the assessment. Furthermore, this dimension claims to be aware of factors which may influence students' performance.

The findings of this study point out that performance is not absolute and context free. Rather, the context and individual differences must be considered in order to achieve fairness. In ideal test design, the individual differences in test-takers should

account for most of the difference in the assessment outcomes. However, in performance assessments we also see differences in instructional settings that may affect individual performance. If context factors are contributing to differences in individual's performance, we must pay attention to them because the data on the performance of test-takers is less accurate without considering the context factors. Hence, to ensure fairness in the context of performance-based assessments, functionally equitable support systems should be developed for test-takers.

When we consider fairness, it is critical to examine the support system that surrounds the learners, including the support of the instructors and the peers as well as the level of institutional support. In the context of performance-based assessments, it is not possible to define an absolute set of context expectations that ensure fairness. Based on this study, it suggests providing individualized learning supports based on their needs for creating equitable learning settings.

The fairness dimension is related to other dimensions, including meaningfulness and educational consequence. Test-takers quickly lose motivation to participate in an assessment when they perceive assessment activities as unfair (Shepard, 2008). The case where participants view the assessment as unfair is unintended and negative consequence. It is undesirable for using a performance-based assessment.

### **Usability**

Usability is an important dimension that should be considered when examining the practicality of a performance-based assessment. For large-scale assessments, costs and efficiency are critical to administer an assessment for a large population of test-takers. It is critical to maintain required resources at acceptable levels. With more

performance-intensive assessments, greater attention is typically given to the development of efficient data collection and scoring procedures.

This study showed that the participants' workload, time investment, efforts, and other resources need to be considered to examine this dimension within the local assessment administration context. However, overemphasizing this dimension without careful consideration of the implementation process may lead to a negative perception of the usability of a performance-based assessment. For example, if a test is assigned within too short period of time, students tend to rush without fully participating in the assessment.

As indicated in chapter four, the participants confirmed that significant workload at the beginning of the edTPA adoption and implementation were inevitable. Their understanding of the assessments increased with their increased engagement in assessment activities. This understanding helped participants develop efficient and strategic approaches to implementing the assessment and to build responses to tasks with a deeper understanding of real-world contexts. The findings suggest an association with other framework constructs in this study. As the participants gained a clearer vision of task requirements (transparency), their ability to decide what to include in response and how to organize them increased. Moreover, the participants started to carefully monitor their teaching ability by reviewing what they already knew and what they could do (educational consequences). It is necessary to understand that with an overemphasis on usability, the participants tended to focus on completing the tasks on time. In such cases, they lost learning opportunities during the assessment activities, which may have directly affected their performance and scores.

In conclusion, this discussion emphasizes a balance between usability and other dimensions. Viability may be an alternative description of usability. Viability is defined as the capacity to operate and sustain tasks successfully under certain conditions. With realistic expectations, participants make thoughtful decisions to use the assessment without being overwhelmed by workloads and unintended stress. Educators often express that performance-based assessment would be ideal under a condition of infinite time, no worries about competing workload, and abundant resources. From a practical perspective, educators should take an active role in figuring out best approaches to administer the assessment and to support students within dynamic assessment contexts.

### **Implications**

In this section, I discuss implications of the edTPA implementation and suggest practical strategies to teacher educators. Also, I discuss implications for the investigation of consequential validity which have not been fully considered in evaluating the validity of performance-based assessments.

#### **Implications for teacher education programs**

edTPA is currently widely used, yet is controversial in teacher education. Although many research studies and teacher educators have confirmed its validity and usefulness for educational and formative purposes, there are a number of skeptics. Some teacher educators harshly critique the implementation of edTPA by using the case of New York state, where edTPA was mandated for awarding a teaching license from the state within a very short adoption timeline. The case of New York is a rapid adoption of these policies which did not allow for professional deliberation or for teacher preparation institutions to accommodate the assessment into their existing programs. Yet, this is not

the situation in other states. Each state has created different policies associated with edTPA, making the comparison of edTPA implementation across states difficult. These policies create substantial changes in the system of teacher preparation, and we see resistance to these changes in general, and especially when the adoption and implementation are fast-tracked.

By using a case of Minnesota in this study, this study urges teacher educators' active role to navigate the best practices for their teacher candidates. edTPA has been implemented in Minnesota during the last four years, and an extremely small number of studies have been published (edTPAminnesota.org). The published studies focused on small contexts of assessment implementation. Even though the studies are valuable to understand what happen in a classroom, it has a limited scope to understand a state-wide implementation. These studies are limited in their ability to generalize administration practices across programs and institutions, yet they provided significant evidence to understand how programs and institutions have adopted edTPA differently based on local decisions.

### **Implications for validity research studies**

Many validity studies have been conducted to evaluate the technical aspects of performance-based assessments. The technical aspect of assessment design and adoption is very important to make the case that the assessment is the appropriate tool to use for its intended measurement goals. Yet, over-emphasis on this aspect of assessment limits our investigations of how assessments are used for improving learning and teaching, which, arguably, is an ultimate goal of assessments in educational contexts. A concept of

consequential validity with special attention to learning and integration into the local context should be considered as critical evidence to evaluate assessment validity.

The consequential aspect of validity draws our attention to the aspects of the assessment process that are not the main interests to test developers. Test developers often do not investigate consequences as evidence of validity, since this type of evidence is generated after assessment implementation. However, it is critical to examine how assessments are used and how they serve as an educative tool for both learners and educators—especially in the case of performance assessments. Validity evidence associated with contextual factors are traditionally believed to be irrelevant or to have unintended effects on test scores or results. The framework explored in this study elaborates on the concept of consequential validity and expands our thinking of validity as a process during implementation as well as before and after implementation. To properly apply this framework, multiple stakeholders' using an assessment should be considered as a part of formal requirement in teacher preparation process. Depending on the design of the assessment system, different stakeholders can potentially participate in this assessment process.

### **Limitations to this study**

As with all research studies, this study has a few limitations. First, my subjectivity should be acknowledged. I believe that a performance-based assessment is a strong educational tool that can provide learning opportunities to all participants and users. I started this study with an assumption that edTPA has positive effects on multiple levels of teacher education programs and their institutions. However, it is important to recognize common concerns and issues raised by other researchers in this field. To

manage my subjectivity, I participated in several national and local edTPA conferences as I was developing this study and sought input on my interview protocols and survey questionnaires from several people. These activities helped me gain a more balanced perspective to develop and complete this study.

Additionally, it is important to note that my position could have influenced the data analysis and data report. I tried to minimize this influence by recruiting relatively large numbers of participants for focus group discussions and individual interviews to secure a variety of perspectives and to provide enough data to allow me to seek trends and consistency across the data.

Another limitation is related to the research design of this study. While the mixed methods design is the most popular and straightforward method of triangulation, as it applies multiple data collection methods, collecting and managing large amounts of data from different sources and analyzing the data thoroughly demands a lot of labor to complete this type of research study. For this reason, some researchers suggest carrying out mixed-methods studies by forming a research team comprising members who have qualitative and quantitative expertise. I did not have the capacity to form a research team for this study. Nevertheless, I consulted with many experts who supported this study. For example, I sent the developed survey questionnaires to the developers of edTPA. They helped me clarify and rewrite the questions. Additionally, I asked my colleagues who are currently working with teacher candidates in teacher education programs in Minnesota to review survey questions. This process helped me increase my expertise in both qualitative and quantitative research.

Lastly, the focus group participants and individual interviewees were sampled in one institution. The total number of the participants was 35, which was sufficient to gather rich data on the perspectives of teacher candidates and faculty on the implementation of edTPA. Their perspectives were strongly related to the contexts of their programs and institution. Teacher candidates and faculty in other institutions may have different perspectives and opinions about their edTPA implementation, depending on their program and institution contexts. With the data collected within a single context, I was able to link the program and institutional contexts with the edTPA implementation.

### **Conclusions**

It is almost impossible to find a perfect assessment to measure human traits. However, it is important to improve the quality of the existing measures. Recently, a growing number of performance-based assessments have been developed and used. When we evaluate the validity of these assessments, it is difficult to ignore the context in which they are intended to be used and the multiple purposes (formative, summative, and evaluative) assumed within one assessment. While the current standards of testing are not ready to accept more flexible approaches to validate performance-based assessments, it is extremely important to evaluate performance-based assessments considering their special characteristics; when this type of assessment is locally implemented in a specific context, context plays a significant role to support learning and to capture accurate performance of students.

A set of six dimensions in this framework contributed to specify complex circumstances of assessment process and its consequences. As I briefly reiterate the set of six dimensions, *Educational consequences* reveal that intended learning outcomes may



be varied depending on assessment contexts. Also, this dimension raised a question how to deal with unintended or negative consequences. *Meaningfulness* shows that a recognition of values of engaging in the assessment activities and processes is related to situated contexts including peers, teachers, existing curriculum and policy, and others. *Directness* is rarely discussed among participants. This dimension is not fully developed enough to stand solely. I view that this dimension could be a part of another dimension, educational consequences. Providing proper feedback and understanding test scores accurately are important part of instruction. An integration of directness with educational consequences would contribute to understand educators' instruction in systematic approaches. *Transparency* re-emphasizes the importance of full understanding of assessment process and goals from all participants. *Fairness* raises a conversation between equity and equality in terms of supporting test-takers who participate in an assessment. Lastly, *usability* may change to *viability* which refers the capacity to operate and sustain tasks successfully under certain conditions. The alternative descriptor guides us to have realistic and practical mindset for engaging in a performance-based assessment.

This framework of consequential validity aims to expand our scope of the assessment implementation in real-world contexts and for educational purposes. This framework embraces the concept of consequential validity to describe what happens within the local context when adopting new performance assessments. By using such a framework, I think that local contexts could be used to better predict the many dimensions of assessment implementation and potentially improve the interpretation and

use of assessment scores and results. A framework such as this would support educators by providing empirical guidance in the implementation of performance-based assessment.

## Bibliography

- American Educational Research Association, American Psychological Association, National Council on Measurement and Education. (1984). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Educational Research Association, American Psychological Association, & National Council on Measurement and Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Educational Research Association, American Psychological Association, & National Council on Measurement and Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.
- Bond, L. (1997, March). Adverse impact and teacher certification. Paper presented at the *Annual Meeting of the American Educational Research Association*, Chicago, IL.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly*, 35(1), 7-28.
- Collins, K. M., Schutz, A. M., & Moss, P. A. (1997). *INTASC candidate interviews: A summary report*. Unpublished manuscript, Interstate New Teacher Assessment and Support Consortium. Washington, DC.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L. & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori and C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 209-240). Thousand Oaks, CA: Sage.
- Darling-Hammond, L., & Adamson, F. (2010, April). *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning*. Stanford Center for Opportunity Policy in Education (SCOPE): Research Brief. Retrieved from [https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary\\_0.pdf](https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary_0.pdf)
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and teacher education*, 16(5), 523-545.

- Darling-Hammond, L., Wise, A. E., & Klein, S. P. (1999). *A license to teach*. San Francisco, CA: Jossey-Bass.
- Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27(4), 307-329. edTPAminnesota.org. (n.d). *Voices from the field*. Retrieved from <http://edtpaminnesota.org/edtpa-in-minnesota/>
- Frederickson, J & Collins, A 1989. A systems approach to educational testing. *Educational Researcher*, 18 (9), 27-32.
- Green, J. (1998). Authentic assessment: Constructing the way forward for all students. *Education Canada*, 38(3), 8-12.
- Greene, J. C., and Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed method evaluation. In J. C. Greene and V. J. Caracelli (eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. New Directions for Evaluation* (pp. 5-17). San Francisco, CA: Jossey-Bass.
- Haertel, E. H. (1991). Performance tests, simulations and other methods. In J. Millman & L. Darling Hammonds (Eds.), *The new handbook of teacher evaluation* (pp. 278–294). Thousand Oaks, CA: Sage.
- Haertel, E. H. (1991). Performance Assessment and Education Reform. *The Phi Delta Kappan*. (80)9, 662-666.
- Hambleton, R. K. (1996, April). Guidelines for Adapting Educational and Psychological Tests. Paper presented at *the Annual Meeting of the National Council on Measurement in Education*. New York, NY.
- Jaeger, R. M. (1996). *Conclusions on the technical measurement quality of the 1995-96 operational version of the National Board for Professional Teaching Standards' middle childhood/generalist assessment*. Unpublished manuscript, National Board for Professional Teaching Standards, Southfield, MI.
- Kennedy, M. (2010). *Teacher assessment and the quest for teacher quality: A handbook*. San Francisco, CA: John Wiley & Sons.
- Linn, R. L. Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*. 20(81), 5-21.
- Livingston, S. (1993). *Inter-assessor consistency of the Praxis III classroom performance assessment: Spring 1992 preliminary version*. Unpublished report, Educational Testing Service, Princeton, NJ.

- McDowell, L. (1995). The impact of innovative assessment on student learning. *Programmed Learning*, 32(4), 302–313.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, 18(2), 5-11.
- Messick, S. (1990). *Validity of Test Interpretation and Use* (Report No. RR-90-11). Educational Testing Service, Princeton, NJ. Retrieved from Educational Testing Website: <http://files.eric.ed.gov/fulltext/ED395031.pdf>
- Messick, S. (1995a). Standards of validity and the validity of standards in performance assessment. *Educational measurement: Issues and practice*, 14(4), 5-8.
- Messick, S. (1995b). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741-749.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Morse, J. M. (1996). Is qualitative research complete? *Qualitative Health Research*, 6, 3-5.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.
- Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational researcher*, 25(1), 20-29.
- Moss, P. A. (1998a). Rethinking validity in the assessment of teaching. In N. Lyons & G. Grant, (Eds.), *With portfolio in hand: Portfolios in teaching and teacher education* (pp. 202-219). New York, NY: Teachers College Press.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of research in education*, 109-162.
- Namey, E., Guest, G., Thairu, L., & Johnson, L. (2008). Data reduction techniques for large qualitative data sets. *Handbook for team-based qualitative research*, 2, 137-161.
- Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002). Two decades of developments in qualitative inquiry a personal, experiential perspective. *Qualitative Social Work*, 1(3), 261-283.

- Pecheone, R. L. and Chung, R. R. (2006). Evidence in Teacher Education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education*, 57(1), 22-36.
- Penk, C., Pöhlmann, C., & Roppelt, A. (2014). The role of test-taking motivation for students' performance in low-stakes assessments: an investigation of school-track-specific differences. *Large-scale Assessments in Education*, 2(5), 1-17.
- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12-15.
- Porter, A., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and their uses. In V. Richardson (Ed.), *Handbook of research on teaching* (pp.259-297). Washington, DC: American Educational Research Association
- Sandelowski, M. (2000). Focus on research methods combining qualitative and quantitative sampling, data collection, and analysis techniques. *Research in nursing & health*, 23, 246-255.
- Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669-700.
- Stanford Center for Assessment, Learning, and Equity (SCALE). (n.d.). *edTPA*. Retrieved from <https://scale.stanford.edu/teaching/edtpa>
- Stanford Center for Assessment, Learning, and Equity (SCALE). (2013, November). *2013 edTPA Field Test: Summary Report*. Retrieved from [https://secure.aacte.org/apps/rl/res\\_get.php?fid=827&ref=edtpa](https://secure.aacte.org/apps/rl/res_get.php?fid=827&ref=edtpa)
- Sweet, D. (1993). *Performance Assessment*. Office of Research, Office of Educational Research and Improvement (OERI) of the U.S. Department of Education. Retrieved from Education Research Consumer Guide: <http://www2.ed.gov/pubs/OR/ConsumerGuides/perfasse.html>
- Shepard, L. A. (1993). Chapter 9: Evaluating Test Validity. *Review of research in education*, 19(1), 405-450.
- Tashakkori, A., & Teddlie, C. (1998). Mixed methodology: *Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Uhlenbeck, A., Verloop, N., & Beijaard, D. (2002). Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment. *The Teachers College Record*, 104(2), 242-272.

- Wei, R. C., & Pecheone, R. L. (2010). Assessment for learning in preservice teacher education. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 69-125). San Francisco, CA: John Wiley & Sons.
- Wise, L. L. (2011, February). Picking up the pieces: Aggregating results from through-course assessments. Paper presented at the *Invitational Research Symposium on Through-Course Summative Assessment*. Atlanta, GA.

## Appendix A

### A List of All Institutions and Those Selected for This Study

List of Institutions in Minnesota		Selected Institutions by Type
1	Metropolitan State University*	Minnesota State Colleges and University System
2	Minnesota State University, Mankato*	
3	Minnesota State University, Moorhead*	
4	North Central University	
5	St. Cloud State University*	
6	Bemidji State University	
7	Southwest Minnesota State University	
8	Winona State University	
9	University of Minnesota, Twin Cities*	University of Minnesota System
10	University of Minnesota, Duluth*	
11	University of Minnesota, Crookston*	
12	University of Minnesota, Morris*	
13	Augsburg College*	Private Institutions
14	Bethel University*	
15	Concordia University, St. Paul*	
16	Hamline University*	
17	Martin Luther College*	
18	St. Catherine University*	
19	University of St. Thomas*	
20	Alfred Adler Graduate School	
21	Bethany Lutheran College	
22	Capella University	
23	Carleton College	
24	College of St. Benedict/St. John's University	
25	College of St. Scholastica	
26	Concordia College/Moorhead	
27	Crown College	
28	Gustavus Adolphus College	
29	St. Mary's University	
30	St. Olaf College	
31	University of Northwestern – St. Paul	
32	Walden University	

\* Selected institutions for surveys.



## **Appendix B**

### **Focus Group Interview Questions for Teacher Candidates**

#### **Introduction**

Good morning and welcome to session. Thanks for taking the time to join us to participate in a focus group in the experiences related to the edTPA. My name is Su Jung Kim, and I am a Ph.D student in curriculum and instruction at the University of Minnesota.

This focus group is part of my dissertation project to understand how the edTPA implemented and how it improve teacher candidate' teaching. There are very few studies available to investigate the implementation process and its effects on teacher candidates' teaching.

You are invited because you have completed the edTPA in your program area recently, so you are the best group of people who can describe the implementation process and make suggestions based on your experiences.

There are no wrong answers but rather differing points of view. Please feel free to share your point of view even if it differs from what others have said. Keep in mind that we're just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.

You've probably noticed the microphone. Please note that this session will be recorded to ensure we adequately capture your ideas during the conversation. However, the comments from the focus group will remain confidential and your name will not be attached to any comments you name. Do you have any questions before we begin?

## CONSEQUENTIAL VALIDITY OF PERFORMANCE-BASED ASSESSMENT

### Potential questions

1. Well, let's begin. I've placed name cards on the table in front of you to help us remember each other's names. Let's find out some more about each other by going around the table. Tell us your name and your program area. If you already know where you are going to teach coming Fall, please share with us.

2. First, think back over last few weeks that you had worked on the edTPA. Many practitioners and researchers argue that the edTPA helps to develop teacher candidates' readiness to teach. What do you think the impact of the edTPA

#### Probe

- a. In what ways was the edTPA helpful to you *improve* your teaching?
- b. What features of the edTPA make you *feel confident* to prepare teaching?
- c. How did edTPA reflections help you to *make decisions* about your teaching?
- d. In what ways do you feel that the edTPA *fell short* in helping you to improve your teaching?
- e. What did you learn about student learning?
- f. What do you think about you know or can do now became your engaged in the edTPA as process?

3. At this point, let's talk about your program supports to work on the edTPA. Now think over the last few months that your program has helped you to prepare/complete the edTPA. Do you feel that you have strong supports to prepare the edTPA?

*Make sure to make a note: this study focus on the process, not just completion.*

#### Probe

- a. You may consider any *course* materials, assignments, classroom activity, guest speakers, and edTPA seminar?
- b. What kind of supports did faculty, supervisors, and cooperatives provide?
- c. Who did you work with (supported you) the most to develop/complete your edTPA?

4. I would like to know how to make our teacher education programs more supportive for completing the edTPA and improving your teaching, and want to hear your thoughts on how we could that.

#### Probe

- a. What your suggestions for teacher education programs could offer to make it supportive for teacher candidates to complete the edTPA successfully?

5. Is there anything else we haven't discussed about your experience with the edTPA?

## Appendix C

### Focus Group Interview Questions for Teacher Educators

#### Introduction

Good morning and welcome to session. Thanks for taking the time to join us to participate in a focus group in the experiences related to the edTPA. My name is Su Jung Kim, and I am a Ph.D student in curriculum and instruction at the University of Minnesota.

This focus group is part of my dissertation project to understand how the edTPA implemented and how it im teacher candidate' teaching. There are very few studies available to investigate the implementation process and its effects on program changes and teacher candidates' teaching.

You are invited because you have worked with teacher candidates closely for completing the edTPA, so you are the best group of people who can teach me the implementation process and make suggestions based on your experiences.

There are no wrong answers but rather differing points of view. Please feel free to share your point of view even if it differs from what others have said. Keep in mind that we're just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.

You've probably noticed the microphone. Please note that this session will be recorded to ensure we adequately capture your ideas during the conversation. However, the comments from the focus group will remain confidential and your name will not be attached to any comments you name. Do you have any questions before we begin?

## CONSEQUENTIAL VALIDITY OF PERFORMANCE-BASED ASSESSMENT

### Potential Questions

1. Tell us your name and your position in your program area. Also, how you have involved in the edTPA implementation.
2. First, think back about the very first moment that you are first introduced to the edTPA?
  - a. What discussions were initiated among teacher educators?
3. Did you make significant changes in your lesson/program level/institutional level?
  - a. What kinds of changes were made?
  - b. How these changes were decided?
  - c. What went particularly well? Or not well?
4. How did you review previous years of teacher candidates' edTPA scores?
  - a. Did you work by yourself? Collaborative with others?
  - b. What lessons did you have?
5. How would you like to see any changes in your classroom, program, and institutions for future implementation?
  - a. These could be actual or potential changes to three different levels